

Rated Tags as a Service – Konzept und Evaluierung

Dissertation

zur Erlangung des akademischen Grades Doktoringenieur (Dr.-Ing.)

Vorgelegt an der
Technischen Universität Dresden
Fakultät Informatik

Eingereicht von

Daniel Kailer (M.Sc.)
geboren am 21. September 1979 in Erding

Gutachter:

Prof. Dr. rer. nat. habil. Dr. h. c. Alexander Schill
Prof. Dr. rer. nat. Jörg Hähner

TU Dresden
Universität Augsburg

Ort und Datum der Verteidigung:
Dresden, den 22. Januar 2016

Kurzfassung

Durch die wachsende Bedeutung des Onlinehandels und der Zunahme an Benutzer-generierten Inhalten werden neue Ansätze benötigt, um Konsumenten bei ihrer Entscheidungsfindung zu unterstützen. Wie Studien zeigen, werden im Onlinehandel häufig Kundenrezensionen und Gesamtbewertungen eingesetzt. Allerdings sind diese beiden Werkzeuge für die Entscheidungsfindung von Konsumenten nur begrenzt hilfreich. Gesamtbewertungen zeigen zwar eine oberflächliche Zufriedenheit der Kunden, geben jedoch keine Auskunft über die Bewertung bestimmter Produktaspekte, z.B. den Tragekomfort von Kopfhörern. Diese Aspekte werden von Kunden häufig in Rezensionen beschrieben, welche jedoch aufgrund ihrer unstrukturierten Weise nicht automatisiert aufbereitet werden können. Konsumenten sind daher gezwungen Rezensionen zu lesen und die darin diskutierten Merkmale manuell zu extrahieren.

Die vorliegende Arbeit leistet mehrere Beiträge zur Adressierung des oben genannten Problems und beschäftigt sich dabei mit der Konzeptionierung, Evaluierung und Dienst-orientierten Bereitstellung einer interaktiven Entscheidungshilfe für den E-Commerce. Zunächst wird anhand einer empirischen Untersuchung der umsatzstärksten Onlineshops aus Deutschland der aktuelle Einsatz von Social Media Features analysiert. Dabei zeigt sich, dass die o.g. Problematik von keinem untersuchten Onlineshop adressiert wird. Ein weiterer Beitrag ist der Entwurf sowie die prototypische Implementierung einer interaktiven Entscheidungshilfe mit der Bezeichnung *Rated Tags*. Rated Tags erlaubt die Benutzer-generierte Definition von bewertbaren Schlagwörtern (Tags) und kombiniert dabei Methoden aus den Bereichen Social Tagging und Bewertungssysteme. Eine nachfolgende Evaluierung des Konzepts im Rahmen einer Anwenderstudie zeigt, dass der Einsatz von Rated Tags die Entscheidungsqualität verbessern sowie den Entscheidungsaufwand von Konsumenten reduzieren kann. Zur Optimierung des Lösungsansatzes wird dann ein Ensemble-Klassifikator aus dem Bereich des überwachten Lernens zur semiautomatisierten Vereinheitlichung von semantisch ähnlichen Tags entworfen, prototypisch implementiert und evaluiert. Die Ergebnisse der Evaluierung zeigen, dass die Leistung des Klassifikators den aktuellen Stand der Technik übersteigt. Als Abschluss der Arbeit wird ein Modell mit der Bezeichnung *Rated Tags as a Service* vorgestellt, welches die Service-orientierte Bereitstellung des Rated Tags-Ansatzes für Onlineshops oder Bewertungsportale beschreibt.

Inhaltsverzeichnis

1	Einleitung	9
1.1	Motivation	9
1.2	Zielsetzung	11
1.3	Aufbau der Arbeit	12
2	Entscheidungsunterstützung im E-Commerce	15
2.1	E-Commerce	15
2.1.1	Definition und Klassifikation	15
2.1.2	Aufbau von E-Commerce Systemen	16
2.2	Entscheidungsfindung	18
2.2.1	Definition	18
2.2.2	Entscheidungskomplexität	19
2.2.3	Entscheidungsstrategien	21
2.3	Entscheidungsunterstützungssysteme	21
2.3.1	Definition und Abgrenzung	21
2.3.2	Passive Entscheidungshilfen	23
2.3.3	Interaktive Entscheidungshilfen	25
2.4	Zusammenfassung	27
3	Social Media im E-Commerce	29
3.1	Social Media	29
3.1.1	Definition und Abgrenzung	29
3.1.2	Klassifikation	31
3.2	Social Commerce	32
3.3	Studie zur Nutzung von Social Media in B2C-Onlineshops	33
3.3.1	Empirischer Forschungsstand	33
3.3.2	Methodisches Vorgehen	35
3.3.3	Ergebnisse	37
3.3.4	Diskussion	40
3.4	Online-Bewertungen	42
3.4.1	Produktbewertungen	43
3.4.2	Produktrezensionen	44
3.5	Tagging	47
3.5.1	Grundlagen	47
3.5.2	Chancen und Risiken	50
3.6	Zusammenfassung	52

4	Konzeptionelle Betrachtung von Rated Tags	53
4.1	Probleme aktueller Bewertungssysteme	53
4.2	Eigener Lösungsansatz	54
4.2.1	Ziel	56
4.2.2	Beispielszenario	56
4.2.3	Akteure	57
4.2.4	Anwendungsfälle	59
4.2.5	Entitäten	64
4.2.6	Herausforderungen	67
4.3	Abgrenzung zu verwandten Ansätzen	68
4.3.1	Lösungsansätze aus der Praxis	68
4.3.2	Verwandte Forschungsarbeiten	69
5	Entwurf und Umsetzung des Rated Tags-Ansatzes	75
5.1	Systemarchitektur	75
5.1.1	Kontextsicht	75
5.1.2	Verteilungssicht	76
5.1.3	Bausteinsicht	78
5.1.4	Laufzeitsicht	83
5.2	Entwurf der Benutzeroberfläche	86
5.3	Prototypische Implementierung	88
5.3.1	Funktionsauswahl	89
5.3.2	Konstruktion	89
6	Anwenderstudie zur Evaluierung des Rated Tags-Ansatzes	95
6.1	Methodisches Vorgehen	95
6.1.1	Datenbeschaffung und Anpassung des Prototyps	96
6.1.2	Erzeugung von Rated Tags	97
6.1.3	Bestimmung des dominierenden Produkts	98
6.1.4	Aufstellung der Hypothesen	100
6.1.5	Auswahl geeigneter statistischer Testverfahren	101
6.1.6	Akquise der Studienteilnehmer	103
6.2	Ergebnisse	106
6.2.1	Manipulation Check	106
6.2.2	Überprüfung der Hypothesen	107
6.2.3	Ergebnisse zur Erstellung von Rated Tags	111
6.2.4	Ergebnisse des Fragebogens	112
6.3	Diskussion	115
6.3.1	Interpretation der Ergebnisse	115
6.3.2	Limitationen der Studie	116
6.3.3	Implikationen für die Forschung	117
6.3.4	Implikationen für die Praxis	118

7	Vereinheitlichung der erstellten Tags	121
7.1	Problemanalyse und Zielsetzung	121
7.2	Verwandte Arbeiten	123
7.2.1	Einordnung	123
7.2.2	Abgrenzung	125
7.3	Klassifikationsmodell	130
7.3.1	Modellübersicht	130
7.3.2	Vorverarbeitung der Daten	133
7.3.3	Hauptklassifikatoren	136
7.3.4	String-basierte Klassifikatoren	145
7.3.5	Ensemble-Klassifikator	152
7.4	Evaluierung	153
7.4.1	Methodisches Vorgehen	154
7.4.2	Ergebnisse	159
7.4.3	Diskussion	167
8	Service-orientierte Bereitstellung von Rated Tags	171
8.1	Zielsetzung und Motivation	171
8.2	Einordnung	172
8.3	Dienstarchitektur	173
8.3.1	Bereitstellungsmodell	173
8.3.2	Datenhaltung	175
8.3.3	Schnittstellenkommunikation	179
8.3.4	Diensteinbindung	184
8.4	Dienstnehmer-übergreifende Nutzung aller Tags	185
9	Zusammenfassung	189
9.1	Ergebnisse der Arbeit	189
9.2	Ausblick	191
	Literaturverzeichnis	195
	Abbildungsverzeichnis	211
	Tabellenverzeichnis	215
	Quelltextverzeichnis	217
	Abkürzungsverzeichnis	219
A	Untersuchung deutscher B2C-Onlineshops	221
B	Evaluierung des Rated Tags-Ansatzes	223

Kapitel 1

Einleitung

1.1 Motivation

Die Bedeutung des Onlinehandels hat in den vergangenen Jahren stark zugenommen. Wie eine Studie aus dem Jahr 2012 zeigt, ist das Onlineshopping besonders in Deutschland sehr beliebt [18]. Laut dieser Studie kaufen etwa zwei Drittel (65 %) der Bundesbürger im Internet ein. Dies liegt deutlich über dem europaweiten Durchschnitt von 45 %. Die steigende Bedeutung des Onlinehandels in Deutschland wird auch am Wachstum der Umsatzzahlen sichtbar. Wie in Abbildung 1.1 dargestellt ist, stieg der Anteil des Onlinehandels am gesamten Einzelhandelsumsatz in den letzten Jahren stetig an.

Ein Grund für das Wachstum des Onlinehandels ist mitunter die einfache Beschaffung von Produktinformationen [155, S. 1]. Onlinehändler versuchen den Kunden möglichst viele Produktinformationen zu präsentieren, da beim Onlineshopping die Produkte, im Gegensatz zum Einzelhandel, von den Konsumenten nicht physisch begutachtet werden können und auch keinerlei Beratungspersonal für Rückfragen zur Verfügung steht. Ferner kann die Informationsmenge seitens des Anbieters nicht eingeschränkt werden, da aufgrund der Anonymität in Onlineshops die speziellen Anforderungen des Konsumenten nicht bekannt sind [155, S. 1].

Neben den vom Onlinehändler bereitgestellten objektiven Informationen sind in Onlineshops auch immer häufiger *Benutzer-generierte Inhalte* (engl. *User-Generated Content, UGC*) anzutreffen. Es ist also auch im Onlinehandel der Trend erkennbar, dass Internetnutzer nicht nur Inhalte konsumieren, sondern auch selbst produzieren. Ein Beispiel aus dem Onlinehandel ist die steigende Beliebtheit zur Erzeugung von Online-Bewertungen für Produkte und Dienstleistungen [17]. Vor allem die in den Rezensionen anderer Kunden beschriebenen Erfahrungen spielen für die Entscheidungsfindung von Konsumenten eine wichtige Rolle [110].

Wie die vorigen Absätze deutlich machen, werden die Besucher eines Onlineshops mit einer Vielzahl an Informationen konfrontiert. Dies ist zwar zum einen positiv, da eine fundierte Basis an objektiven und subjektiven Produktinformationen zur Verfügung steht. Zum anderen besteht jedoch bei einer steigenden Menge an Informationen auch die Gefahr einer Informationsüberflutung (engl. *information overload*). Wie Chewning und Harrell beschreiben, reduziert sich die Entscheidungsleistung einer Person ab einer gewissen Informationsmenge [30]. Dies trifft vor allem auf die Rezensionen innerhalb von Onlineshops zu, welche eine Vielzahl an Produktinformationen enthalten, jedoch aufgrund ihrer unstrukturierten Natur nicht einfach analysiert werden können und daher von den

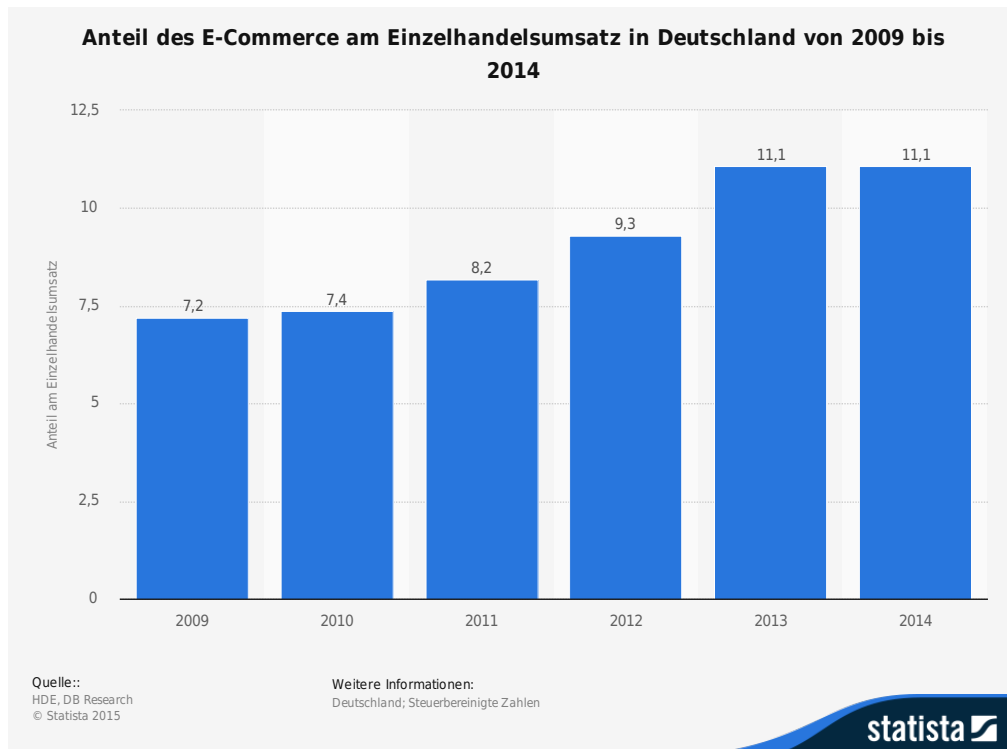


Abbildung 1.1: Anteil des E-Commerce am Einzelhandelsumsatz in Deutschland (Quelle: [184])

Konsumenten einen hohen kognitiven Aufwand erfordern. Die gestiegene Entscheidungskomplexität führt gemäß Johnson und Payne oft dazu, dass Konsumenten im Austausch für einen reduzierten Entscheidungsaufwand eine geringere Entscheidungsqualität in Kauf nehmen [87].

Aufgrund der beschriebenen Problematik sind in Onlineshops spezielle Entscheidungshilfen erforderlich, welche Konsumenten in ihrer Entscheidungsfindung unterstützen. Ariely schlägt hierzu die Einführung von interaktiven Informationssystemen vor, damit Anwender die Informationsmenge basierend auf ihren persönlichen Anforderungen einschränken können [8, S. 233]. Die Berücksichtigung der spezifischen Konsumentenansforderungen wird in aktuellen Bewertungssystemen jedoch nicht ausreichend unterstützt. Die zur Verfügung stehenden Gesamtbewertungen helfen Konsumenten nicht in der Überprüfung ihrer speziellen Anforderungen, sondern dienen nur als allgemeiner Qualitätsindikator. Benutzer-generierte Rezensionen adressieren hingegen zwar oft eine Vielzahl an Produktmerkmalen, jedoch ist es für den Konsumenten sehr zeitaufwendig die darin diskutierten Merkmale zu analysieren und zu vergleichen. Das Lesen aller Rezensionen für alle infrage kommenden Alternativen ist auch meist gar nicht möglich bzw. zielführend. Eine weitere Schwachstelle aktueller Systeme ist, dass Konsumenten vor dem Lesen einer Rezension nicht ermitteln können, ob die jeweilige Rezension im Hinblick auf die Anforderungen überhaupt relevant ist.

1.2 Zielsetzung

Die Problematik aus dem vorigen Abschnitt hat deutlich gemacht, dass im Onlinehandel neue Entscheidungshilfen benötigt werden, welche die Schwachstellen aktueller Bewertungssysteme ausgleichen. Ein Ziel dieser Arbeit ist zunächst die Erfassung des Stands der Technik in Hinblick auf die Verwendung von Social Media in deutschen Onlineshops. Dazu soll eine empirische Untersuchung durchgeführt werden, welche Aufschluss darüber geben soll, wie Social Media momentan eingesetzt wird und an welchen Stellen es als Grundlage für eine neue Entscheidungshilfe verwendet werden könnte.

Basierend darauf soll eine neue interaktive Entscheidungshilfe konzipiert werden, welche auf Social Media basiert und insbesondere die zuvor genannten Probleme aktueller Bewertungssysteme adressiert. Da es sich bei der Entscheidungshilfe um eine hybride Form von Benutzer-generierten Schlagwörtern (engl. *tags*) und Kundenbewertungen handelt, wird für das vorgestellte Konzept die Bezeichnung *Rated Tags* verwendet. Ein wichtiges Ziel von Rated Tags ist es, die von Vig et al. genannte Lücke zwischen Gesamtbewertungen und Rezensionen zu schließen [200, S. 323]. Durch die entworfene Entscheidungshilfe soll vor allem der kognitive Aufwand für Konsumenten reduziert werden, was gemäß Hu et al. gerade bei Einkaufsentscheidungen im Onlinehandel von großer Bedeutung ist [77, S. 42]. Um die Praxistauglichkeit der entworfenen Entscheidungshilfe zu erproben, soll basierend auf dem Entwurfskonzept eine prototypische Implementierung sowie eine entsprechende Evaluierung des Ansatzes durchgeführt werden. Zudem soll ein Konzept zur Vereinheitlichung der erstellten Tags vorgestellt und evaluiert werden, um den Aufwand für Konsumenten weiter zu reduzieren.

Ein weiteres Ziel der Arbeit ist die Erarbeitung eines Konzepts zur Service-orientierten Bereitstellung der entworfenen Entscheidungshilfe. Der Dienst soll eine möglichst einfache Einbindung der Entscheidungshilfe in bestehende Bewertungsportale oder Onlineshops ermöglichen. Die angestrebte Zielgruppe der Dienstanutzer sind dabei vor allem kleine und mittlere Unternehmen (KMU). Die Service-orientierte Einbindung der Entscheidungshilfe bedeutet für KMU einen geringeren Entwicklungs- und Administrationsaufwand als bei einer Eigenentwicklung. Zudem werden die der Entscheidungshilfe zugrunde liegenden Inhalte von den Konsumenten selbst produziert, wodurch nur ein geringer personeller Aufwand seitens der KMU notwendig wird.

Basierend auf der oben genannten Zielsetzung, lassen sich die relevanten Themengebiete der Arbeit wie folgt identifizieren. Eine zentrale Rolle spielt der Themenkomplex E-Commerce. Hier ist vor allem der Stand der Technik in Bezug auf Entscheidungsunterstützungssysteme, sowohl aus wissenschaftlicher wie auch aus praktischer Sicht, relevant. Ein wichtiger Berührungspunkt sind zudem Benutzer-generierte Inhalte, wie z.B. Social Tagging oder Bewertungssysteme. Für eine teilautomatische Vereinheitlichung von Tags werden zudem Techniken aus dem überwachten maschinellen Lernen verwendet. Für den konzeptuellen Entwurf des Modells in Service-orientierter Weise spielen ferner Themenkomplexe wie Service-orientierung und Software as a Service (SaaS) eine wichtige Rolle. Als übergeordneter Rahmen dieser Arbeit kann der Themenkomplex der verteilten Systeme angesehen werden, da dieser sowohl die Grundlage für die Bereiche E-Commerce und Social Media als auch für die speziellen Service-orientierten Konzepte bildet.

1.3 Aufbau der Arbeit

Im Anschluss an dieses Kapitel erfolgt zunächst die Aufbereitung der relevanten Grundlagen für diese Arbeit sowie eine kritische Betrachtung von verwandten Forschungsarbeiten. Dabei beleuchtet Kapitel 2 zunächst die Grundlagen aus den Bereichen E-Commerce, Entscheidungsfindung und Entscheidungsunterstützungssysteme. Ein wichtiges Ziel dieses Kapitels ist die Definition und Abgrenzung von Entscheidungshilfen im E-Commerce. Das darauffolgende Kapitel 3 betrachtet die Grundlagen aus dem Themenkomplex Social Media sowie dessen aktuellen Einsatz im E-Commerce. Das Kapitel umfasst unter anderem die Erfassung des aktuellen Stands der Technik anhand einer empirischen Untersuchung in Hinblick auf die Nutzung von Social Media in deutschen Onlineshops. Die für diese Arbeit relevanten Social Media-Ausprägungen „Online-Bewertungen“ und „Tagging“ werden dann detaillierter betrachtet.

In Kapitel 4 wird das Forschungsproblem anhand eines Beispielszenarios verdeutlicht. Danach erfolgt die konzeptionelle Betrachtung des eigenen Lösungsansatzes, welcher im Rahmen dieser Arbeit unter der Bezeichnung *Rated Tags* geführt wird. Für den genannten Lösungsansatz werden die für das Modell relevanten Akteure, Anwendungsfälle und Entitäten sowie potenzielle Risiken näher betrachtet. Der Lösungsansatz wird dann sowohl aus praktischer als auch aus wissenschaftlicher Sicht zu den jeweiligen verwandten Arbeiten abgegrenzt.

Basierend auf der konzeptionellen Betrachtung widmet sich Kapitel 5 der prototypischen Umsetzung von *Rated Tags*. Dabei wird im Entwurf insbesondere auf die Systemarchitektur und auf die Anforderungen an die Benutzeroberfläche der Entscheidungshilfe eingegangen. Im Rahmen der prototypischen Implementierung werden dann die zentralen Aspekte des Entwurfs identifiziert und in Form einer Web-basierten Anwendung umgesetzt.

Anhand des in Kapitel 5 entwickelten Prototyps wird in Kapitel 6 eine Anwenderstudie zur Evaluierung des Konzepts durchgeführt. Betrachtet wird hier primär die Reduktion des zeitlichen und kognitiven Aufwands durch die Entscheidungshilfe. Das Kapitel beschreibt dabei zunächst das methodische Vorgehen für die Gestaltung und Durchführung der Studie. Ferner werden in Hinsicht auf die entworfene Entscheidungshilfe drei Hypothesen aufgestellt, welche im Rahmen der Studie überprüft werden. Es folgt eine Interpretation der Studienergebnisse sowie eine Betrachtung der Limitationen der Studie. Abschließend werden die Implikationen der Studie sowohl für die Praxis als auch für die weitere Forschung diskutiert.

Um die Qualität des *Rated Tags*-Systems zu verbessern, wird in Kapitel 7 ein Konzept zur teilautomatisierten Vereinheitlichung von Benutzer-generierten Tags vorgestellt. Dies ist sinnvoll, da die Gruppierung von semantisch ähnlichen Tags den kognitiven Aufwand des Anwenders reduzieren kann. Für diese Vereinheitlichung werden unter anderem Methoden des überwachten maschinellen Lernens verwendet. Es wird ein Ensemble-Klassifikator konzipiert, welcher mehrere heterogene Klassifikatoren miteinander vereint. Ferner folgen eine prototypische Umsetzung sowie eine Evaluierung des Klassifikators. Die Methoden und Ergebnisse des entworfenen Klassifikators werden dann mit dem aktuellen Stand der Technik verglichen.

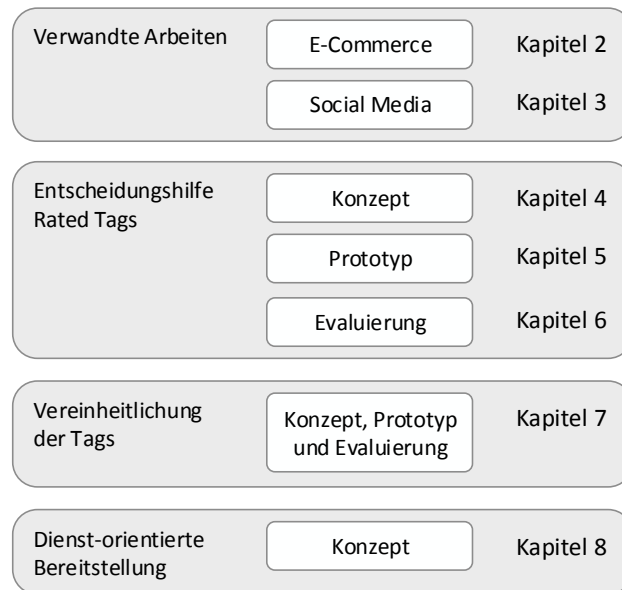


Abbildung 1.2: Übersicht der Kernkapitel dieser Arbeit

In Kapitel 8 wird ein Konzept vorgestellt, um den entworfenen und evaluierten Ansatz in Service-orientierter Weise anderen Bewertungsportalen oder Onlineshops zur Verfügung zu stellen. Dabei werden unter anderem die Aspekte Datenhaltung, Kommunikationsschnittellen und Einbindung des Dienstes näher betrachtet. Zudem wird ein Modell vorgestellt, um Synergieeffekte aus der Dienstnehmer-übergreifenden Nutzung der existierenden Tags zu erzielen.

Abschließend werden die Ergebnisse der Arbeit in Kapitel 9 zusammengefasst. Dabei werden zunächst die wissenschaftlichen Beiträge der Arbeit diskutiert. Danach erfolgt ein Ausblick auf zukünftige Verwertungs- und Erweiterungsmöglichkeiten des vorgestellten Konzepts. Eine grafische Übersicht zu den Kernkapiteln dieser Arbeit ist in Abbildung 1.2 dargestellt.

Kapitel 2

Entscheidungsunterstützung im E-Commerce

In diesem Kapitel wird der Begriff *E-Commerce* definiert und die für diese Arbeit relevanten Ausprägungen näher erläutert. In diesem Zusammenhang werden auch die typischen Komponenten einer E-Commerce Architektur analysiert. Ferner werden die Grundlagen der Entscheidungsfindung im E-Commerce sowie existierende entscheidungsunterstützende Hilfen näher betrachtet. Abschließend wird das Kapitel in Abschnitt 2.4 zusammengefasst und in Bezug zur Forschungsarbeit gesetzt.

2.1 E-Commerce

2.1.1 Definition und Klassifikation

Die wörtliche Übersetzung des Begriffes *Electronic Commerce* (kurz *E-Commerce*) steht für *elektronischen Handel*. Der Begriff E-Commerce steht also im engeren Sinn für die elektronische Handelsunterstützung beim Kauf und Verkauf von Produkten oder Dienstleistungen [113, 137]. Weitere Definitionen des Begriffes fokussieren allerdings nicht nur auf eine elektronische Unterstützung, sondern im Speziellen auf eine Handelsunterstützung durch das Internet oder Intranet [114, 197]. Kann der Großteil der Transaktionen digital abgewickelt werden, so werden im Deutschen auch synonym die Begriffe *Internethandel* oder *Onlinehandel* verwendet [205].

Eine eindeutige Definition des E-Commerce ist jedoch in der existierenden Literatur nicht vorzufinden. So definieren Kotler und Bliemel den Begriff E-Commerce als „die Abwicklung von Kundenbetreuung und Geschäftsprozessen über elektronische Medien“, wobei hier auch Transaktionen innerhalb von Organisation explizit erfasst werden [103]. Diese etwas breiter gefasste Definition wird von anderen Autoren wiederum nicht als E-Commerce bezeichnet, sondern unter dem Schlagwort *Electronic Business* (kurz *E-Business*) subsumiert [113, 114]. In Anlehnung an Laudon et al. wird der Begriff E-Business für diese Arbeit als Oberkategorie des E-Commerce eingestuft [113]. Für die weitere Arbeit wird jedoch nur der enger gefasste Begriff des E-Commerce eine Rolle spielen.

Wie in den vorigen Absätzen erläutert wurde, ist die elektronische Unterstützung für den Kauf und Verkauf von Produkten oder Dienstleistungen die Hauptcharakteristik des E-Commerce. Für eine weitere Einordnung des Begriffes können die Transaktionsteilnehmer sowie die Umsetzung der Handelsunterstützung klassifiziert werden.

Die bedeutendsten Teilnehmer des E-Commerce sind Unternehmen (engl. *businesses*), Konsumenten bzw. Endkunden (engl. *consumers*) und öffentliche Einrichtungen [113, 137]. Letzterer Teilnehmertyp spielt im Rahmen dieser Arbeit keine Rolle und wird daher nicht näher betrachtet. Die relevanten Segmente ergeben sich somit durch die Kombination der Teilnehmergruppen Business und Consumer. Diese Segmente des E-Commerce werden nachfolgend näher erläutert [113, S. 575]:

- *Business-to-Business (B2B)*
Umfasst alle Transaktionen die zwischen Unternehmen durchgeführt werden.
- *Business-to-Consumer (B2C)*
Umfasst den Verkauf von Produkten oder Dienstleistungen von Unternehmen an Konsumenten. Ein bekannter Vertreter dieses Segments ist der Onlineshop von Amazon¹.
- *Consumer-to-Consumer (C2C)*
Umfasst alle Transaktionen die zwischen Endkunden durchgeführt werden. Ein bekanntes Beispiel aus diesem Segment ist das Online-Auktionshaus Ebay².

Für den Rahmen dieser Arbeit ist nur die B2C-Ausprägung des E-Commerce interessant, d.h. der elektronische Handel zwischen Unternehmen und Konsumenten. Alle nachfolgenden Verwendungen des Begriffes E-Commerce zielen daher auf die B2C-Ausprägung des E-Commerce ab.

Eine weitere Klassifizierung des E-Commerce lässt sich jedoch nicht nur anhand der Teilnehmertypen vornehmen, sondern auch anhand der bereitgestellten bzw. genutzten Infrastruktur. Ein immer populärer werdender Bereich ist *Mobile Commerce*. Als Mobile Commerce werden kommerzielle Transaktionen bezeichnet, die über mobile Endgeräte, wie z.B. Tablets oder Smartphones, abgewickelt werden [114, S. 59]. Werden innerhalb des E-Commerce soziale Interaktionsmöglichkeiten seitens des Anbieters bereitgestellt, so spricht man hier von dem Bereich des *Social Commerce* [112]. Diese spezielle Form des E-Commerce spielt in dieser Arbeit eine zentrale Rolle und wird daher in Abschnitt 3.2 noch detaillierter betrachtet.

2.1.2 Aufbau von E-Commerce Systemen

Die Literaturrecherche im Bereich E-Commerce hat ergeben, dass kein anerkanntes Referenzmodell zum Aufbau von E-Commerce Systemen existiert, welches alle Kundenanforderungen erfüllt [46]. Nichtsdestotrotz gibt es essenzielle Komponenten in typischen E-Commerce Systemen. Betrachtet man den Aufbau einer E-Commerce Anwendung auf einer detaillierteren Ebene, so entspricht die Architektur typischerweise dem Client-Server Paradigma. Die Client- und Server-seitigen Aspekte einer E-Commerce Anwendung werden nachfolgend kurz erläutert.

¹<http://www.amazon.com> (besucht am 01.04.2014)

²<http://www.ebay.com> (besucht am 01.04.2014)

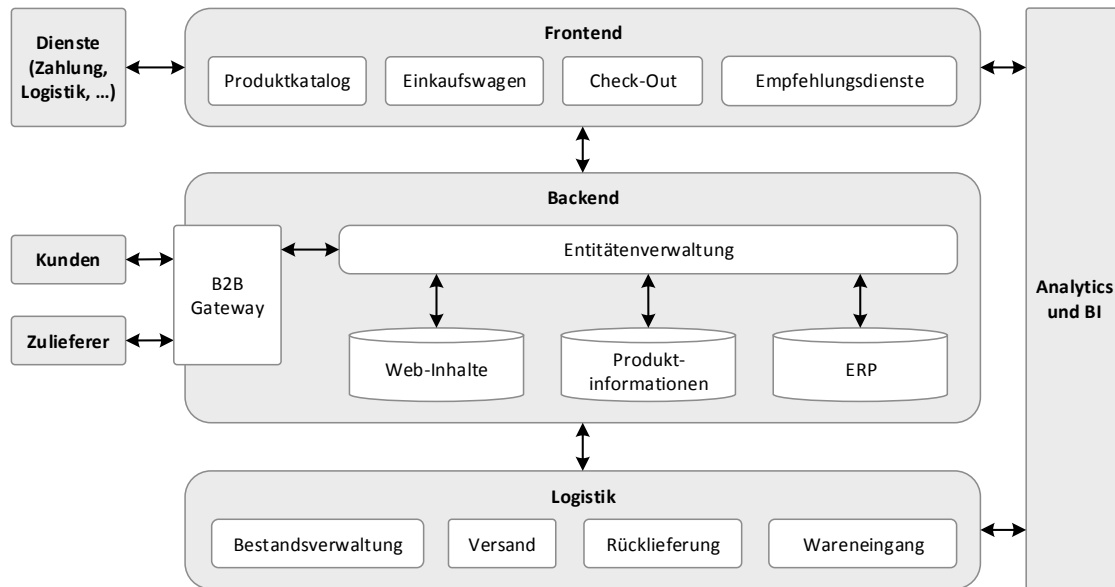


Abbildung 2.1: Beispielhafte Architektur eines Onlineshop-Systems auf hoher Ebene (leichte Modifikation von [109, S. 551])

Server-seitig

In Abbildung 2.1 wird beispielhaft eine Server-seitige Systemarchitektur eines Onlineshop-Systems auf hoher Ebene aufgezeigt. Für den Rahmen dieser Arbeit sind primär die dort dargestellten Komponenten *Frontend* und *Backend* interessant. Die Backend-Komponente umfasst unter anderem die Datenhaltung für die Web-Inhalte, Produktinformationen sowie eine mögliche Anbindung an ein Enterprise-Resource-Planning (ERP)-System. Die Frontend-Komponente stellt hingegen die für die Konsumenten nutzbaren Dienste, wie beispielsweise eine Warenkorb-Funktionalität oder Empfehlungsdienste, bereit. Eine tiefer gehende Analyse von detaillierten Onlineshopping-Funktionen ist in der Arbeit von Treutner und Ostermann zu finden. Die Autoren zeigen die Evolution von Funktionen und Elementen von Onlineshops basierend auf praktischen und wissenschaftlichen Ansätzen [193].

Um die Anfragen der Clients zu verarbeiten, sind neben dem Webserver Server-seitig häufig noch weitere Server im Einsatz, z.B. Applikationsserver oder Datenbankserver. Bei dieser klassischen Verteilung findet der Hauptteil der Nachrichtenverarbeitung auf dem Server statt, wodurch dieser zum Engpass wird [10]. Gerade im E-Commerce sind jedoch kurze Antwortzeiten von hoher Bedeutung, da viele Konsumenten bei längeren Ladezeiten den Seitenaufbau abbrechen und womöglich bei einem konkurrierenden Onlineshop einkaufen [52, 63]. Nicht nur aus diesem Grund ist in Onlineshops immer häufiger eine *Rich Internet Application (RIA)* anzutreffen. Bei einer RIA übernimmt der Anwendungscient einen größeren Teil der Verarbeitung als bei einer klassischen Web-Architektur [10]. Der nachfolgende Abschnitt betrachtet kurz zwei mögliche Client-seitige Implementierungsmöglichkeiten.

Client-seitig

In traditionellen E-Commerce Anwendungen wird der Client durch den Webbrowser des Anwenders dargestellt. Dieser kommuniziert mittels *Hypertext Transfer Protocol (HTTP)* mit dem Webserver des jeweiligen Onlineshops und erhält ein *Hypertext Markup Language (HTML)*-Dokument zurück, welches im Webbrowser dargestellt wird. Für die Umsetzung einer RIA spielen hier besonders Technologien wie *JavaScript* oder *Asynchronous JavaScript and XML (AJAX)* eine Rolle, wobei letztere einen asynchronen Datenaustausch und somit eine bessere Reaktionszeit der Anwendung ermöglicht. Die typischen Datenformate für die Übertragung basieren häufig auf *eXtensible Markup Language (XML)* oder *JavaScript Object Notation (JSON)*.

In den letzten Jahren hat sich die Nutzung von mobilen Endgeräten (wie beispielsweise Smartphones und Tablets) stark gesteigert. Dieser Trend ist auch im E-Commerce spürbar, da Konsumenten diese mobilen Endgeräte auch immer öfter für Onlineeinkäufe nutzen. Hillman et al. definieren die Nutzung von mobilen Endgeräten für den Onlineeinkauf als *Mobile Commerce* oder kurz *M-Commerce* [68, S. 113]. Auf den mobilen Endgeräten steht den Anwendern typischerweise auch ein Webbrowser zur Verfügung, der für das Onlineshopping genutzt werden kann. Allerdings unterliegen die mobilen Endgeräte gewissen Besonderheiten (z.B. einem relativ kleinem Display). Daher sind auf diesen Geräten oft native mobile Anwendungen zu finden, welche für die jeweiligen Geräte optimiert sind. Diese Anwendungen stellen neben den Webbrowsern eine weitere Client-seitige Umsetzung für E-Commerce bzw. M-Commerce Systeme dar.

2.2 Entscheidungsfindung

Der in dieser Arbeit vorgestellte Dienst soll den Entscheidungsfindungsprozess für Konsumenten des E-Commerce verbessern. In diesem Abschnitt werden zunächst die Grundlagen der Entscheidungsfindung erläutert. Ferner wird die im E-Commerce vorhandene Entscheidungskomplexität aufgrund der Vielzahl an Produktalternativen näher betrachtet. Abschließend werden die für diese Arbeit wichtigsten Entscheidungsstrategien kurz diskutiert.

2.2.1 Definition

In Anlehnung an Turban beschreibt der Begriff *Entscheidungsfindung* einen Prozess, bei dem ein Entscheidungsträger aus mehreren Alternativen genau eine Alternative auswählt, um ein definiertes Ziel zu erreichen [196]. Der Ablauf eines solchen Entscheidungsfindungsprozesses, auch in Bezug auf den E-Commerce, wurde bereits in diversen Arbeiten betrachtet [98, 104, 115, 140]. Für diese Arbeit wird der Entscheidungsprozess, in Anlehnung an Kotler und Keller sowie Laux und Liermann, in die folgenden Phasen untergliedert [104, S. 166] [115, S. 33]:

1. Definition des Problems (und Präzisierung des Zielsystems)
2. Informationssuche

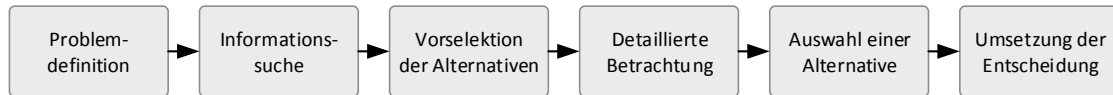


Abbildung 2.2: Phasen eines Entscheidungsprozesses im E-Commerce (in Anlehnung an [62, 104, 115])

3. Analyse der möglichen Alternativen
4. Auswahl einer Alternative
5. Umsetzung der Entscheidung für die ausgewählte Alternative

Zur Verdeutlichung der oben genannten Phasen wird nachfolgend ein beispielhafter Entscheidungsprozess im Bereich des B2C-E-Commerce dargestellt. Eine Person möchte eine DVD abspielen, besitzt jedoch keinen passenden DVD-Player. Der DVD-Player sollte auch über einen USB-Anschluss verfügen, um Bilder und Videos von externen USB-Geräten abspielen zu können (Phase 1). Die Person besucht einen Onlineshop und sucht basierend auf seinen Anforderungen nach einem passenden DVD-Player (Phase 2). Die Person studiert Kundenbewertungen und vergleicht die gefundenen Alternativen (Phase 3). Nachdem die Alternativen hinsichtlich der Anforderungen überprüft und gegenübergestellt wurden, entscheidet sich die Person für ein Produkt (Phase 4). Zur endgültigen Umsetzung der Entscheidung fügt die Person das ausgewählte Produkt seinem Warenkorb hinzu und beendet nach Angabe der Liefer- und Zahlungsmodalitäten die Transaktion (Phase 5).

Für den Rahmen dieser Arbeit sind besonders die Phasen 2 und 3 des o.g. Entscheidungsprozesses von Bedeutung, also die Informationssuche und die Analyse von Alternativen. Diese beiden Phasen stellen besonders im E-Commerce eine Herausforderung für Konsumenten dar, da im Vergleich zum stationären Handel typischerweise viel mehr Alternativen zur Auswahl stehen. Laut Häubl und Trifts kann die Phase zur Analyse der Alternativen in zwei weitere Unterphasen eingeteilt werden, wobei je nach Stadium die Tiefe der Informationsverarbeitung variiert [62]. In der ersten Phase werden die Alternativen nur oberflächlich betrachtet und auf die erfolgsversprechendsten Alternativen reduziert. Die Untermenge der für relevant gehaltenen Alternativen wird auch als *Consideration Set* bezeichnet [195, S. 89]. Diese erste Phase kann auch als Vorselektion bezeichnet werden. In der nachfolgenden Phase werden die selektierten Alternativen dann detaillierter überprüft. Ersetzt man die von Laux und Liermann beschriebene Analysephase (Phase 3) durch die von Häubl und Trifts definierten Unterphasen, so ergibt sich der in Abbildung 2.2 dargestellte Entscheidungsprozess für den Bereich des E-Commerce.

2.2.2 Entscheidungskomplexität

Anbieter im E-Commerce stellen ihren Besuchern üblicherweise eine Vielzahl an Produkten zur Verfügung. Dies hängt unter anderem auch damit zusammen, dass ein positiver Zusammenhang zwischen der Produktvielfalt und der Kundenzufriedenheit identifiziert

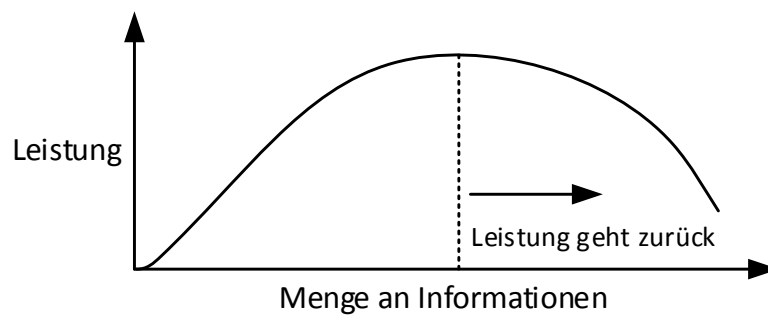


Abbildung 2.3: Abnahme der Entscheidungsleistung ab einer gewissen Menge an Informationen

wurde [70, S. 527]. Jedoch steigt durch die im E-Commerce vorhandene Vielfalt an Alternativen auch der kognitive Aufwand bei den Entscheidungsträgern, um all diese Alternativen zu vergleichen. Gemäß Mitchell et al. und Shugan ist diese Überfülle an Informationen (engl. *information overload*) nicht nur durch die Anzahl der Alternativen bedingt, sondern auch durch die entscheidungsrelevanten Informationen (Attribute) der Alternativen selbst [136, S. 143] [179]. Mit dem Begriff *Attribut* ist hier das Merkmal einer Alternative gemeint, z.B. die Akkulaufzeit eines Smartphones. Diese Komplexität hängt jedoch immer von der jeweiligen Produktkategorie ab. So existieren bei einem Smartphone beispielsweise relativ viele entscheidungsrelevante Attribute, z.B. die Akkulaufzeit, die Größe des Displays, das Vorhandensein einer integrierten Kamera sowie deren Qualität und einige mehr. Diese Kriterien müssen dann für alle Alternativen überprüft und verglichen werden.

Mit Zunahme der bereitgestellten Informationen wird üblicherweise auch die Entscheidungsqualität erhöht, jedoch nur bis zu einem gewissen Punkt. Ab diesem Punkt führt eine weitere Zuführung von neuen Informationen zu einem Abfall der Entscheidungsqualität [30, S. 539]. Dieser Verlauf kann grafisch als umgekehrtes *U* dargestellt werden (siehe Abbildung 2.3). Ein zusätzlich festgestelltes Phänomen ist, dass Personen ab einer gewissen Menge an Informationen dazu tendieren die Entscheidung aufzuschieben [210, 290f.]. Für den E-Commerce bedeutet dies, dass Besucher bei zu vielen unverarbeiteten Informationen den Einkauf abbrechen und eventuell bei einem anderen Onlineshop einkaufen.

Bei Entscheidungen mit steigender Komplexität hat sich gezeigt, dass Konsumenten für einen reduzierten Entscheidungsaufwand eine Reduzierung der Entscheidungsgenauigkeit in Kauf nehmen [14, 87]. Dies bedeutet, dass sich Entscheidungsträger oft mit einer zufriedenstellenden, aber nicht optimalen, Alternative begnügen. Diese Konstellation ist besonders dann anzutreffen, wenn die Anzahl an Alternativen sehr hoch ist bzw. wenn ein Vergleich dieser Alternativen sehr schwierig ist, d.h., wenn die Komplexität der Entscheidung zunimmt [154]. Die Erleichterung von komplexen Entscheidungen ist Aufgabe von *Entscheidungsunterstützungssystemen*. Diese werden in Abschnitt 2.3 noch näher betrachtet.

Die Komplexität einer Entscheidung ergibt sich also durch die Anzahl der Alternativen und die Anzahl der entscheidungsrelevanten Attribute. Wie umfangreich die Evaluation der Alternativen und derer Attribute ausfällt, hängt von der jeweiligen Strategie des Entscheiders ab. Der nachfolgende Abschnitt soll daher kurz die Grundlagen und Varianten von Entscheidungsstrategien verdeutlichen. Entscheidungsstrategien stellen laut Pfeiffer den Stand der Technik für die Beschreibung des Kaufverhaltens von Konsumenten dar [155, S. 18].

2.2.3 Entscheidungsstrategien

Eine Entscheidungsstrategie ist eine Abfolge von Handlungen, um von einem initialen Wissensstand zu einem Zielzustand zu gelangen, in dem der Entscheidungsträger das Entscheidungsproblem als gelöst betrachtet [153, S. 109]. Im Kontext dieser Arbeit beschreibt eine Entscheidungsstrategie den Prozess, um sich aus einer Vielzahl an Alternativen für eine Alternative zu entscheiden. Die Charakteristiken der bestehenden Entscheidungsstrategien sind für die Arbeit relevant, da sie die mögliche Vorgehensweise von Konsumenten bei ihrer Entscheidungsfindung beschreiben. Dies wiederum bildet die Grundlage für Entscheidungsunterstützungssysteme, welche in Abschnitt 2.3 näher erläutert werden.

In der bestehenden Literatur lassen sich eine Vielzahl von Entscheidungsstrategien finden. So enthält die Literaturanalyse von Pfeiffer den Vergleich von fünfzehn unterschiedlichen Entscheidungsstrategien [155], welche wiederum auf unterschiedlichen Charakteristiken basieren [167]. Für den Rahmen dieser Arbeit sind im Speziellen nur zwei Entscheidungsstrategien relevant, welche basierend auf Pfeiffer nachfolgend kurz betrachtet werden [155, S. 20–22].

Bei der ersten Strategie handelt es sich um die konjunktive Strategie (engl. *conjunctive strategy*). Bei dieser Strategie werden diejenigen Alternativen ausgeschlossen, für die der Wert eines Attributs unter einem definierten Grenzwert (engl. *threshold*) liegt, z.B. Digitalkameras die eine Mindestanzahl an Megapixel nicht erfüllen. Der Konsument überprüft bei dieser Strategie die für ihn relevanten Attribute und übernimmt nur Alternativen, deren Attributwerte über dem Grenzwert liegen.

Die zweite für diese Arbeit relevante Strategie basiert auf einer lexikographischen Heuristik. Hierzu gewichtet der Entscheidungsträger die relevanten Attribute nach ihrer Wichtigkeit. Das für den Entscheider wichtigste Attribut wird dann bei allen Alternativen überprüft. Ergibt sich bei der Überprüfung eine dominierende Alternative hinsichtlich des wichtigsten Attributs, so wird diese Alternative ausgewählt. Ist dies nicht möglich, z.B. bei Gleichheit zweier Alternativen, so wird mit dem zweitwichtigsten Attribut fortgefahren und so weiter.

2.3 Entscheidungsunterstützungssysteme

2.3.1 Definition und Abgrenzung

Wie die vorigen Abschnitte gezeigt haben, führt die Zunahme von Alternativen und deren Informationen zu einer Erhöhung der Entscheidungskomplexität. Zur Reduzierung

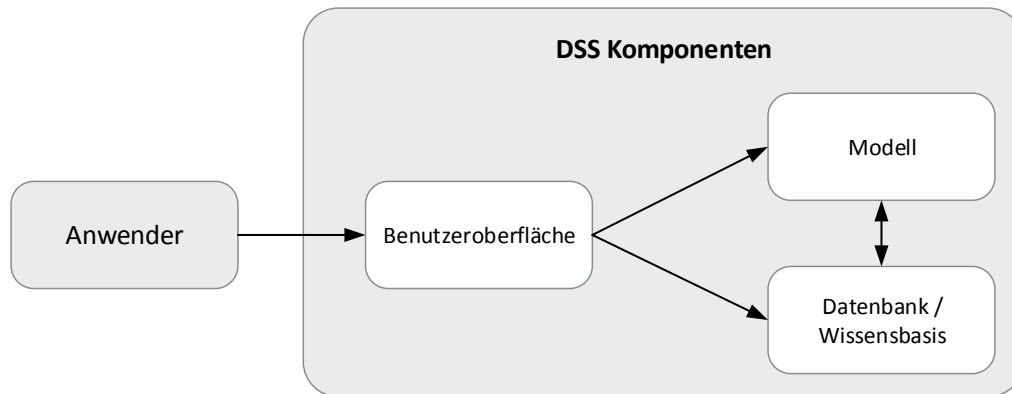


Abbildung 2.4: Komponenten eines DSS (in Anlehnung an [7, S. 1047])

dieser Komplexität und damit zu einer Erhöhung der Entscheidungsqualität hat sich in der Vergangenheit ein *Entscheidungsunterstützungssystem (EUS)* (engl. *Decision Support System, DSS*) etabliert. Nachfolgend wird nur die Abkürzung DSS verwendet, da diese in der Literatur häufiger anzutreffen ist als die Abkürzung EUS. Ein DSS ist ein Informationssystem, das zur Unterstützung von komplexen Entscheidungen oder zur Problemlösung verwendet wird [177, S. 111]. Es soll einen Entscheidungsträger nicht ersetzen, sondern anhand von Daten und einer klaren Benutzeroberfläche die Entscheidung erleichtern [176, S. 279].

In Anlehnung an Ariav und Ginzberg umfassen die drei Hauptkomponenten eines DSS die Benutzeroberfläche, das Modell sowie die dahinter liegende Wissensbasis (siehe Abbildung 2.4). Diese Komponenten lassen sich basierend auf Ariav und Ginzberg wie folgt abgrenzen [7]: Die Benutzeroberfläche ist für die Ein- und Ausgabe der entscheidungsrelevanten Information zuständig. Die Wissensbasis (meist in Form einer Datenbank) enthält die gesamten (entscheidungsrelevanten) Daten des jeweiligen Informationssystems. Die Modell-Komponente umfasst spezielle Algorithmen oder Abläufe, um die in der Wissensbasis vorhandenen Informationen zu aggregieren oder in übersichtlicher Art und Weise aufzubereiten. Wie in Abbildung 2.4 dargestellt ist, interagiert der Anwender mit der Benutzeroberfläche, welche einen Zugriff auf die beiden anderen Komponenten erlaubt.

In der Literatur lassen sich verschiedene Typen von DSS abgrenzen [71, 161], wobei für den Rahmen dieser Arbeit nur die Unterkategorie der Web-basierten DSS relevant ist. Bei einem Web-basierten DSS handelt es sich um ein DSS, welches über das Web bereitgestellt und vom Anwender über einen Webrowsers bedient wird [161, 218]. Shim et al. grenzen die Zielgruppe eines Web-basierten DSS auf Manager oder Business Analysten ein [177]. Eine Literaturanalyse zu DSS hat ergeben, dass als Zielgruppe eines DSS häufig Manager genannt werden [113, 142, 161, 177, 214]. Da diese Arbeit jedoch die Entscheidungsunterstützung von Konsumenten betrachtet, erscheint der explizite Bezug zum Begriff DSS als unpassend. Passender ist die englische Bezeichnung *Customer Decision Support System (CDSS)*, welche auch von anderen Autoren in diesem Kontext verwendet wird [113, 138, 142, 155].

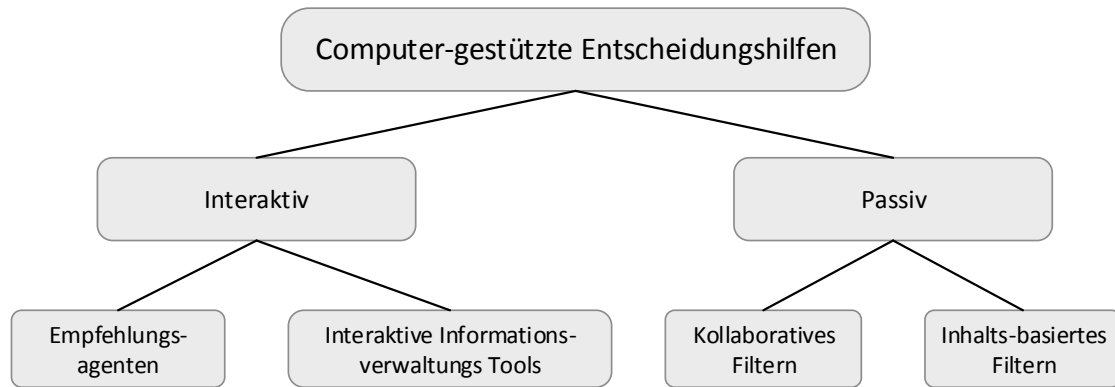


Abbildung 2.5: Kategorisierung von Entscheidungshilfen (übersetzte und modifizierte Version von [155, S. 95])

Im Bereich des CDSS beschreiben Häubl und Trifts informationsverarbeitende Anwendungen als *Computer-gestützte Entscheidungshilfen* (engl. *computer-aided decision aids*), die sie als Teil eines CDSS ansehen [62, S. 3]. Der Begriff Computer-gestützte Entscheidungshilfe (nachfolgend nur als Entscheidungshilfe bezeichnet) wird ebenfalls in weiteren wissenschaftlichen Beiträgen verwendet, die sich speziell mit der Entscheidungsunterstützung von Konsumenten im E-Commerce beschäftigen [62, 155, 191]. Daher wird im Rahmen dieser Arbeit nicht mehr von CDSS, sondern von Entscheidungshilfen die Rede sein, wobei mit Entscheidungshilfen hier Computer-gestützte Entscheidungshilfen im Bereich des E-Commerce gemeint sind.

Häubl und Trifts definieren Entscheidungshilfen als standardisierbare, informationsverarbeitende Anwendungen, die durch ein Computer-basiertes System ausgeführt werden, um den kognitiven Aufwand des Entscheidungsträgers zu reduzieren [62]. Konsumentenstudien haben gezeigt, dass Entscheidungshilfen im E-Commerce die Entscheidungsqualität steigern sowie den Entscheidungsaufwand minimieren können [13, 62, 201], wobei dies jedoch nicht immer der Fall sein muss [13, S. 208]. Entscheidungshilfen können grob in interaktive und passive Entscheidungshilfen unterteilt werden [22]. Ein entsprechendes Kategorisierungsschema ist in Abbildung 2.5 dargestellt. Es zeigt die Aufteilung in interaktive und passive Entscheidungshilfen, wobei die Unterknoten dieser Kategorien nicht erschöpfend dargestellt sind. Die nachfolgenden Abschnitte werden die interaktiven und passiven Entscheidungshilfen voneinander abgrenzen und beispielhaft erläutern.

2.3.2 Passive Entscheidungshilfen

Entscheidungsunterstützende Anwendungen ohne aktive Benutzerbeteiligung werden als *passive Entscheidungshilfen* bezeichnet [22, S. 226]. Die bekanntesten Vertreter dieser Art sind *Empfehlungssysteme* (engl. *recommender systems*). Ein Empfehlungssystem empfiehlt einem Benutzer aus einer Menge an Entitäten eine Untermenge, welche anhand des Benutzerkontextes für sinnvoll erachtet wird [100, 164]. Für diesen Abschnitt der passiven Entscheidungshilfen wird von automatisierten Empfehlungsdiensten ausgegangen,

welche keine explizite Interaktion mit dem Benutzer erfordern [182]. Im Bereich des E-Commerce schlagen Empfehlungsdienste beispielsweise Produkte vor, die dem Kunden, basierend auf vergangenen Einkäufen oder sonstigen Kontext-sensitiven Informationen, gefallen könnten [168, S. 158].

Empfehlungsdienste verwenden verschiedenste Empfehlungsansätze für die Generierung von Vorschlägen. Zwei der üblichsten Ansätze sind das *Kollaborative Filtern* (engl. *collaborative filtering*) sowie das *Inhalts-basierte Filtern* (engl. *content-based filtering*). Diese beiden Ansätze sind ebenfalls als Unterknoten der passiven Entscheidungshilfen in Abbildung 2.5 dargestellt. Ricci et al. klassifizieren noch weitere Empfehlungsansätze [164], welche jedoch im Rahmen dieser Arbeit nicht näher betrachtet werden. Nachfolgend werden nur die Ansätze des Collaborative Filtering und Content-based Filtering erläutert und voneinander abgegrenzt.

Laut Ricci et al. ist Collaborative Filtering der populärste und am meisten verbreitetste Ansatz von Empfehlungssystemen [164]. Wie der Name vermuten lässt, ist dieser Ansatz kollaborativer Natur, d.h., er basiert auf Vergangenheitswerten anderer Anwender mit ähnlichen Präferenzen [164]. Diese Präferenzen werden beispielsweise durch Bewertungen, bereits getätigte Einkäufe oder durch vergangenes Nutzungsverhalten ermittelt. Gibt es beispielsweise einen Kunde K1 der die Produkte P1, P2 und P3 gekauft hat und einen Kunden K2 der die Produkte P1 und P2 gekauft hat, so könnte ein Empfehlungssystem dem Kunden K2 das Produkt P3 vorschlagen, da die Präferenzen der beiden Kunden in P1 und P2 übereinstimmen.

Im Gegensatz zum Collaborative Filtering bezieht das Content-based Filtering nicht die Präferenzen anderer Nutzer mit ein, sondern generiert eine Empfehlung anhand von vergangenen Bewertungen oder Einkäufen des jeweiligen Anwenders [164]. Bei dieser Art der Empfehlung werden die Merkmale der in der Vergangenheit präferierten Produkte analysiert und Produkte mit ähnlichen Merkmalen vorgeschlagen. Hat ein Kunde beispielsweise in der Vergangenheit einen Film mit Jack Nicholson als sehr gut bewertet, so könnte der Inhalts-basierte Ansatz weitere Filme mit identischen Merkmalen vorschlagen, z.B. Filme in denen der Schauspieler Jack Nicholson ebenfalls mitgespielt hat.

Da jeder Empfehlungsansatz gewisse Vor- und Nachteile aufweist, ist oft auch eine Kombination mehrerer Ansätze anzutreffen. Solche *hybriden Empfehlungssysteme* kombinieren verschiedenste Empfehlungsansätze, um die Nachteile einzelner Ansätze zu kompensieren [164]. Welche konkreten Ansätze die heutigen Onlinehändler für ihre Empfehlungen heranziehen ist oft nicht bekannt. Auf der Webseite von Amazon.de werden zu Produkten häufig auch Empfehlungen angepriesen unter dem Titel „*Kunden, die diesen Artikel gekauft haben, kauften auch*“, was auf einen Collaborative Filtering Ansatz hindeuten würde.

Einige Empfehlungssysteme (z.B. das von Amazon.de) erläutern den Anwendern zudem, wieso ein bestimmtes Produkt empfohlen wurde. Es besteht dann mitunter die Möglichkeit für eine Empfehlung anzugeben, ob man mit dieser zufrieden ist oder nicht. Bei dieser Art von Empfehlungssystem findet zwar eine gewisse Einbindung des Anwenders statt, jedoch kann diese in diesem Kontext nur als reaktiv und nicht als interaktiv klassifiziert werden.

2.3.3 Interaktive Entscheidungshilfen

Im Gegensatz zu den passiven Entscheidungshilfen ist bei den interaktiven Entscheidungshilfen (engl. *Interactive Decision Aids, IDAs*) eine Interaktion des Anwenders erforderlich. Steuer definiert Interaktivität als das Ausmaß zu welchem ein Anwender die Form und den Inhalt einer Umgebung in Echtzeit beeinflussen kann [185, S. 84]. Obwohl diese Art der Entscheidungshilfe primär für Konsumenten konzipiert ist, findet sich in der Literatur auch der explizite begriffliche Bezug in der Bezeichnung *Interactive Consumer Decision Aid (ICDA)* wieder [138], wobei im Rahmen dieser Arbeit nur die Abkürzung IDA verwendet wird. IDAs sollen Konsumenten bei einer informierten Einkaufsentscheidung unterstützen und folglich zu einer Erhöhung der Entscheidungsqualität sowie zu einer Reduzierung des Entscheidungsaufwands führen [138] [208, S. 295].

Unglücklicherweise gibt es in der existierenden Literatur keinen Konsens über die Kategorisierung von IDAs. Gupta et al. betrachten IDAs zum einen als Verwaltung der bereitgestellten Informationen (*Interactive Information Management Tools, IIMTs*) und zum anderen als Bereitstellung von Erläuterungen zu Produkten (*Interactive Information Comprehension Tools, IICTs*) [58]. Pfeiffer zählt interaktive Informationsverwaltungstools (IIMTs) ebenfalls zu den IDAs, ersetzt jedoch IICTs durch Empfehlungsagenten (engl. *Recommendation Agents, RAs*) [155]. Wie Abbildung 2.5 zu entnehmen ist, hält sich diese Arbeit an die Kategorisierung von Pfeiffer und kategorisiert IDAs in RAs und IIMTs. Eine Abgrenzung dieser beiden Begriffe wird nachfolgend vorgenommen.

Ein Empfehlungsagent (engl. *Recommendation Agent, RA*) kann als spezielle Ausprägung eines Empfehlungssystems verstanden werden. In Abschnitt 2.3.2 wurde explizit von automatisierten Empfehlungssystemen gesprochen, welche keine Benutzerinteraktion erfordern. In Anlehnung an verwandte Arbeiten kann ein RA als eine Entscheidungshilfe definiert werden, welche einen Benutzer explizit nach dessen Präferenzen fragt [62, 155, 182], z.B. welche Merkmale eines Produkts besonders wichtig sind (Gewichtung von Merkmalen). Im Gegensatz dazu verwenden Xiao und Benbasat den Begriff Empfehlungsagent synonym zum Begriff Empfehlungssystem und unterscheiden daher nicht nach expliziter und impliziter Empfehlungsart [214]. In dieser Arbeit wird diese Unterscheidung jedoch gemacht.

Neben RAs sind IIMTs ein fester Bestandteil von IDAs. Gupta et al. definieren IIMTs als Anwendungen, die Konsumenten eine Sortierung, Filterung und Gegenüberstellung von Alternativen ermöglichen, z.B. anhand von bestimmten Merkmalen [58, S. 163]. Dadurch können diese Anwendungen die Konsumenten in den Phasen der Alternativenanalyse unterstützen, also sowohl in der Vorselektion der Alternativen als auch bei der detaillierten Betrachtung des Consideration Sets [155] (vergleiche hierzu Abbildung 2.2). Laut Gupta et al. wird durch die Unterstützung von IIMTs das Vertrauen der Konsumenten in den Onlinehändler positiv beeinflusst [58]. Eine empirische Evaluation von Pfeiffer hat gezeigt, dass Kunden IIMTs gegenüber RAs bevorzugen [155, S. 93]. Ferner stellen IIMTs laut Pfeiffer die vorherrschende Form von IDAs im Internet dar [155, S. 96].

Die Möglichkeit zur Sortierung von Produktalternativen, z.B. nach Preis oder anhand von Kundenbewertungen, ist in aktuellen Onlineshops weit verbreitet. Diese Möglichkeit unterstützt die in Abschnitt 2.2.3 genannte lexikographische Heuristik. Hat ein Konsument

Ihre Vergleichslisten:
Fernseher/Monitore

		
	€ 399,-*	€ 149,90*
	Auf Lager	Auf Lager
		
Details		
Effizienzklasse		
sichtbares Bild	73,66 cm	61 cm
Auflösung	2560x1080 Pixel	1920x1080 Pixel
Format	21:9	16:9
Paneltechnik	AH-IPS	TN

Abbildung 2.6: Beispielhafte Vergleichstabelle zweier Monitore (Quelle: <http://www.alternate.de/html/product/compare/page.html>, besucht am 01.04.2014)

beispielsweise keine speziellen Anforderungen an ein Produkt, so kann es eine Strategie sein, nur nach dem Preis zu urteilen, d.h., sein wichtigstes Kriterium für die Entscheidung ist der Preis. Über eine aufsteigende Sortierung aller infrage kommenden Alternativen lässt sich dann die günstigste Alternative leicht ermitteln. Das Kriterium Preis ist hier nur beispielhaft zu verstehen. Es könnte auch eine Sortierung eines Produktmerkmals stattfinden, sofern der Onlineshop dies unterstützt.

Über das Filtern von Produktalternativen wird die zweite in Abschnitt 2.2.3 beschriebene Entscheidungsstrategie (Conjunctive Strategy) ermöglicht. Häufig lassen sich bei der Produktsuche in Onlineshops gewisse Grenzbereiche festlegen, um Produkte hinsichtlich von Minimal- oder Maximalwerten zu filtern, z.B. zur Festlegung eines Maximalpreises eines Produkts. Abhängig vom jeweiligen Anbieter ist jedoch auch ein Attribut-basierter Filter denkbar, der dem Anwender das Filtern anhand von Produktmerkmalen erlaubt. Dadurch lassen sich Produktalternativen eliminieren, die ohnehin nicht infrage kommen. Ein Beispiel wäre die Einschränkung der Bilddiagonale oder Auflösung bei Monitoren.

Eine weitere interaktive Entscheidungshilfe ist die *Vergleichstabelle* (engl. *comparison matrix*). In einer Vergleichstabelle werden die (Produkt-)Alternativen typischerweise in Spalten und die Attribute (Produktmerkmale) in Zeilen angezeigt. Dies soll einen schnellen Vergleich der Produkte erlauben. Die eigentliche Interaktivität solcher Anwendungen besteht im Hinzufügen und Entfernen von Alternativen oder Merkmalen. Eine beispielhafte Vergleichstabelle ist in Abbildung 2.6 dargestellt. Sie zeigt zwei Monitore, die hinsichtlich ihrer Produktmerkmale gegenübergestellt werden.

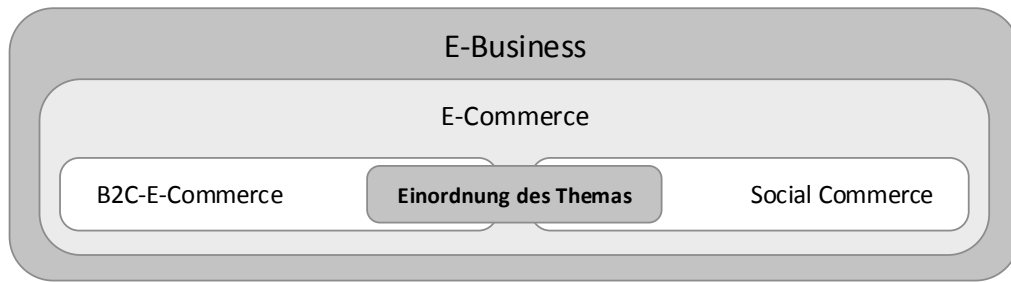


Abbildung 2.7: Einordnung des Themas im Bereich des E-Commerce

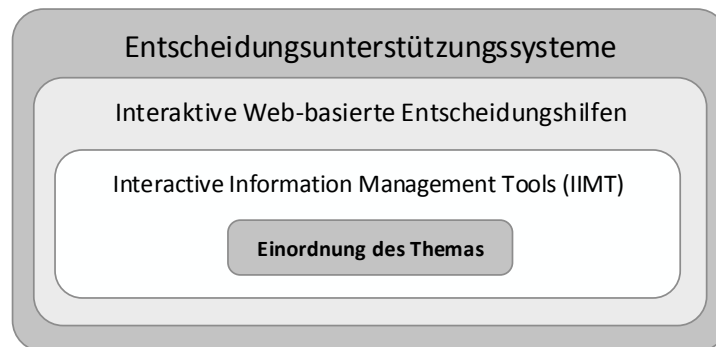


Abbildung 2.8: Einordnung des Themas im Bereich Entscheidungsunterstützung

2.4 Zusammenfassung

Der Bereich des E-Commerce, als Unterkategorie des E-Business, befasst sich mit dem Ein- und Verkauf von Produkten und Dienstleistungen über elektronische Medien. Der Fokus der Arbeit beschäftigt sich primär mit der Web-basierten B2C-Ausprägung des E-Commerce, d.h. mit Web-basierten Transaktionen zwischen Unternehmen und Endkonsumenten. Der zunehmende Einfluss sozialer Medien im Onlinehandel, klassifiziert als Social Commerce, ist ebenfalls ein relevantes Themengebiet der Arbeit. So befindet sich die Arbeit aus Sicht des E-Commerce in der Schnittmenge zwischen B2C-E-Commerce und Social Commerce (siehe Abbildung 2.7).

Wie Abschnitt 2.2.2 gezeigt hat, führt die Überfülle an Informationen, speziell im Onlinehandel bedingt durch die Vielzahl an verfügbaren Produkten und deren Merkmale, bei Konsumenten zu einem erhöhten Aufwand in der Entscheidungsfindung. Es haben sich Entscheidungsunterstützungssysteme etabliert, welche ursprünglich für Manager gedacht waren, jedoch auch immer häufiger für Endanwender eingesetzt werden. Vor allem interaktive Web-basierte Entscheidungshilfen sind im Bereich des E-Commerce sehr verbreitet, um Endkonsumenten bei ihrer Einkaufsentscheidung zu unterstützen. Die Forschungsarbeit positioniert sich speziell im Bereich der interaktiven Informationsverwaltungs-Tools (IIMTs), welche die Entscheidungsfindung durch das Sortieren, Filtern und Vergleichen von Alternativen verbessern sollen. Die Einordnung des Themas für den Bereich der Entscheidungsunterstützung ist in Abbildung 2.8 dargestellt.

Kapitel 3

Social Media im E-Commerce

In diesem Kapitel werden die Grundlagen aus dem Bereich Social Media in Bezug zum E-Commerce aufbereitet. Zunächst folgen eine begriffliche Definition sowie eine Abgrenzung zu verwandten Themengebieten. Danach wird der Begriff des Social Commerce, welcher bereits im vorigen Kapitel kurz angesprochen wurde, definiert und abgegrenzt. Es folgt eine Aufbereitung des Stands der Technik bezüglich des Einsatzes von Social Media in B2C-Onlineshops anhand einer empirisch durchgeführten Studie. Anschließend werden die für diese Arbeit relevanten Themenbereiche des Social Media detailliert betrachtet. Abschließend werden die Ergebnisse des Kapitels zusammengefasst und in Beziehung zur Forschungsarbeit gesetzt.

3.1 Social Media

3.1.1 Definition und Abgrenzung

Die Evolution des Webs ist mit einer stetig steigenden Interaktivität verbunden, bei der Anwender mehr und mehr Informationen aktiv im Web zur Verfügung stellen, während Anwender in den Anfangszeiten des Webs die Inhalte eher passiv konsumiert haben. Diese Veränderung bzw. diese neue Version des Webs wurde 2005 von O'Reilly mit dem Schlagwort *Web 2.0* titulierte [145]. Leider existiert keine exakte Definition des Begriffes Web 2.0, wodurch es in der bestehenden Literatur häufig zu Überschneidungen in der Begriffsdefinition kommt. Ferner ist der Begriff Web 2.0 laut Ebersbach et al. auch kommerziell vorgebelegt und als unscharf zu bezeichnen [45, S. 23]. Trotz alledem ist es wichtig den Begriff Web 2.0 zu definieren, da er in der existierenden Literatur häufig mit dem Begriff *Social Media* in Verbindung gebracht wird. Auch aufgrund der oben genannten Unschärfe versuchte sich O'Reilly 2006 erneut an der Definition des Begriffes Web 2.0:

„Web 2.0 is the business revolution in the computer industry caused by the move to the internet as platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them. (This is what I've elsewhere called „harnessing collective intelligence.“)“ [144]

Aus obiger Definition lässt sich ableiten, dass das Web eine immer größere Rolle als Plattform einnimmt und dass Web-basierte Anwendungen die Kollaboration von

Anwendern unterstützen sollten, um Netzwerkeffekte auszunutzen. Eine Betrachtung des Web 2.0 kann basierend auf dieser Definition von drei verschiedenen Seiten aus vorgenommen werden. Von sozialer Seite aus stehen die Zusammenarbeit zwischen Anwendern und die Erstellung von Benutzer-generierten Inhalten (engl. *User-Generated Content*, *UGC*) im Vordergrund [11]. Aus technischer Sicht sind speziell Technologien relevant, welche für die Entstehung des Web 2.0 wegweisend waren, beispielsweise *AJAX* oder *Really Simple Syndication (RSS)* [9, 11, 96]. Die wirtschaftliche Seite von Web 2.0, welche im Rahmen dieser Arbeit nicht näher betrachtet wird, umfasst mitunter Aspekte der Kundenbindung oder Kundengewinnung [11].

Während in der Literatur des Öfteren die Begriffe Web 2.0 und Social Media synonym verwendet werden (z.B. durch [35, 170]), soll in dieser Arbeit zwischen diesen beiden Begriffen differenziert werden. Eine in der Literatur oft zitierte Definition von Social Media basiert auf Kaplan und Haenlein:

„Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content“ [96, S. 61]

An der Definition von Kaplan und Haenlein ist jedoch eine gewisse Redundanz zu kritisieren, da die ideologischen Grundlagen des Web 2.0 bereits die Erzeugung und den Austausch von Benutzer-generierten Inhalten umfassen. Die zentrale Rolle der Benutzer-generierten Inhalte wird auch in der nachfolgenden Definition von Kietzmann et al. hervorgehoben:

„Social media employ mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content“ [97, S. 241]

Aus den beiden obigen Definitionen zu Social Media lässt sich auch ableiten, dass es sich hier um Software handelt, die eine Kollaboration von Nutzern fördert. Dieser Betrachtungspunkt deckt sich mit dem Schlagwort *Social Software*, welches vor dem Aufkommen des Begriffes Social Media in diesem Kontext betrachtet wurde. Hippner beschreibt Social Software als Web-basierte Anwendungen, die „den Informationsaustausch, den Beziehungsaufbau und die Kommunikation in einem sozialen Kontext unterstützen“ [69, S. 7]. Richter und Koch hingegen definieren Social Software als „Anwendungssysteme, die auf Basis neuer Entwicklungen im Bereich der Internettechnologien und unter Ausnutzung von Netzwerk- und Skaleneffekten, indirekte und direkte zwischenmenschliche Interaktion [...] ermöglichen und die Beziehungen ihrer Nutzer im World Wide Web abbilden und unterstützen.“ [166, S. 8]. Vergleicht man die obigen Definitionen, so können die Begriffe Social Media und Social Software synonym verwendet werden.

Der explizite Bezug auf die Web-basierte Natur von Social Media hat zum Begriff *Social Web* geführt. Dieser Begriff wird ebenfalls synonym zu den Begriffen Social Media und Social Software verwendet [170]. Diese Abgrenzung ist allerdings nicht ganz korrekt, da Social Media nicht nur im Kontext Web-basierter Technologien, sondern auch im mobilen Sektor Anwendung findet.

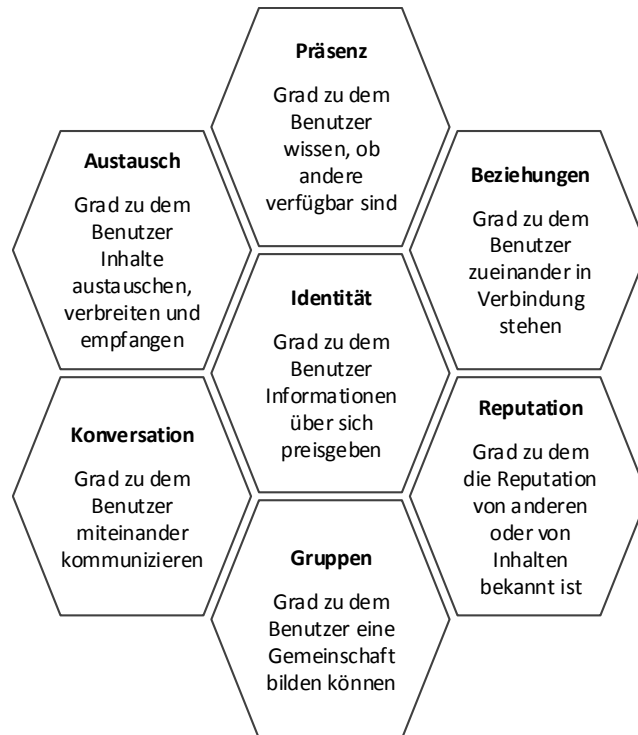


Abbildung 3.1: Bausteine von Social Media (in Anlehnung an [97, S. 243])

3.1.2 Klassifikation

Wie der vorige Abschnitt gezeigt hat, können Social Media als interaktive Internet-basierte Anwendungen klassifiziert werden. Einzelne Ausprägungen dieser Anwendungen werden in der Literatur oft als *Social Media Features* bezeichnet [38, 81]. Ein beispielhaftes Social Media Feature wäre die „Gefällt mir“-Funktion von Facebook, um Inhalte zu bewerten. Wie die Definition aus dem vorigen Abschnitt gezeigt haben, ist der Begriff Social Media, sowie die damit verbundenen Social Media Features, relativ breit gefächert. Um Social Media Features besser klassifizieren zu können, schlagen Kietzmann et al. daher die in Abbildung 3.1 dargestellten Social Media Bausteine vor. Diese Bausteine werden in Anlehnung an Kietzmann et al. nachfolgend kurz erläutert [97]:

Identität Bestimmt den Grad zu dem Benutzer Informationen über sich preisgeben. Dieser Baustein ist bei sozialen Netzwerken (engl. *social networks*) in denen der Benutzer und seine Identität im Vordergrund stehen besonders ausgeprägt, z.B. bei Facebook¹.

Präsenz Bestimmt den Grad zu dem Benutzer wissen, ob andere Benutzer verfügbar sind. Dieser Baustein spielt unter anderem bei kommunikationsorientierten Social Media

¹<http://www.facebook.com> (besucht am 01.04.2014)

Features eine Rolle, z.B. um anwesenden Benutzern eine Nachricht zu schicken. Aber auch Onlinespiele setzen diesen Baustein häufig ein, um ein Zusammenspiel von zwei oder mehreren Spielern zu ermöglichen.

Beziehungen Bestimmt den Grad zu dem Benutzer zueinander in Verbindung stehen. Dieser Baustein ist ebenfalls verstärkt in sozialen Netzwerken vorzufinden. Unterschieden wird hier häufig die Art der Beziehung. Freundschaften auf Facebook sind in der Regel bidirektional, d.h., wenn Person A mit Person B befreundet ist, dann ist auch Person B mit Person A befreundet. Beziehungen auf der Microblogging Plattform Twitter¹ können hingegen auch unidirektionaler Art sein, d.h., ein Anwender kann einem anderen Anwender folgen, ohne dass dieser ihm auch folgen muss.

Reputation Bestimmt den Grad zu dem die Reputation von anderen Benutzern oder Inhalten bekannt ist. Beispiele hierzu gibt es aus sozialen Netzwerken, aber auch aus dem E-Commerce. In einigen sozialen Netzwerken und Onlineshops haben Mitglieder oft die Möglichkeit Inhalte zu bewerten, z.B. anhand eines „Daumen hoch“-Buttons oder anhand einer 5-Sterne Bewertungsskala. Die Reputation ist dann für alle Benutzer sichtbar und kann für weitere Zwecke verwendet werden, z.B. für eine Sortierung nach der Beliebtheit von Inhalten.

Gruppen Bestimmt den Grad zu dem Benutzer in der Lage sind eine Gemeinschaft zu bilden. Dieser Baustein ist unter anderem dann sinnvoll, wenn Kontakte in Gruppen eingeteilt werden sollen, z.B. Freunde, Arbeitskollegen etc.

Konversation Bestimmt den Grad der Kommunikation zwischen den jeweiligen Benutzern. Mögliche Ausprägungen sind beispielsweise Nachrichten-basierte Systeme oder Anwendungen, die eine Echtzeit-Kommunikation in Form eines Chatsystems anbieten.

Austausch Bestimmt den Grad zu dem Benutzer Inhalte generieren und austauschen können. Dieser Baustein bezieht sich also primär auf den Aspekt der Benutzer-generierten Inhalte. Die Art der ausgetauschten Inhalte hängt von der Spezialisierung der jeweiligen Anwendung ab. So existieren beispielsweise spezielle Social Media Features für Bilder oder Videos.

3.2 Social Commerce

In den vergangenen Jahren haben die Konzepte Web 2.0 und Social Media die Entwicklung des E-Commerce maßgeblich beeinflusst [213]. Durch die immer stärker gewordene Interaktivität hat sich der E-Commerce von einer Produkt-orientierten zu einer Kunden-orientierten Umgebung gewandelt [80]. Dieser Wandel bzw. diese neue Ausrichtung wird häufig als *Social Commerce* bezeichnet. Wang und Zhang definieren Social Commerce als „eine Form des E-Commerce, welche durch Social Media angetrieben wird“ [206].

¹<http://www.twitter.com> (besucht am 01.04.2014)

3.3 Studie zur Nutzung von Social Media in B2C-Onlineshops

Marsden und Huang et al. unterteilen Social Commerce in zwei Arten. Zum einen in die Umsetzung von E-Commerce auf Social Media Plattformen (wie z.B. bei dem sozialen Netzwerk Facebook) und zum anderen in die Einbindung von Social Media innerhalb von E-Commerce Plattformen (wie z.B. bei dem Onlinehändler Amazon) [81, 130]. Diese Arbeit wird sich mit letzterem Fall beschäftigen, also mit der Einbindung von Social Media in E-Commerce Plattformen.

Social Commerce ist ein breites, interdisziplinäres Forschungsfeld, in dem verschiedenste Forschungsprobleme behandelt werden, z.B. das Konsumentenverhalten oder diverse Social Network Technologien [220, S. 61]. Im Rahmen dieser Forschungsarbeit stehen besonders die Aspekte Systementwurf und Entscheidungsfindung im Vordergrund. Gemäß Curty und Zhang soll Social Commerce die Entscheidung von Konsumenten unterstützen, indem Informationen nicht nur vom Onlinehändler, sondern auch von anderen Kunden bereitgestellt werden [37]. Bekannte Social Media Features zur Unterstützung der Entscheidungsfindung sind Produktbewertungen und Kundenrezensionen, welche bereits zu frühen Zeiten des E-Commerce angeboten wurden. Jedoch sind mit dem technologischen Fortschritt des Webs und mit der immer weiteren Verbreitung von sozialen Netzwerken auch immer mehr Möglichkeiten zur Interaktion entstanden. Es stellt sich daher die Frage, welche Social Media Features aktuell von Onlinehändlern eingesetzt werden und ob womöglich Potenziale für weitere Social Media Features im E-Commerce existieren. Zur Beantwortung dieser Fragestellung wurde eine empirische Untersuchung von B2C-Onlineshops durchgeführt, welche im nachfolgenden Abschnitt näher erläutert wird.

3.3 Studie zur Nutzung von Social Media in B2C-Onlineshops

Wie bereits erwähnt wurde, existieren schon seit den frühen Zeiten des E-Commerce Interaktionsmuster wie Produktbewertungen oder Kundenrezensionen. Allerdings stellt sich die Frage wie sich die technologische Entwicklung des Webs sowie der immer stärker werdende Anteil an Benutzer-generierten Inhalten auf die Social Media Features im E-Commerce auswirken. Wie nutzen aktuelle Onlinehändler Social Media, um die Konsumenten hinsichtlich ihrer Kaufentscheidung zu unterstützen? Zur Klärung dieser Punkte wurde eine empirische Studie durchgeführt. Dabei wurden insgesamt 115 deutsche B2C-Onlineshops hinsichtlich ihres Einsatzes von Social Media untersucht und klassifiziert. Die nachfolgenden Unterabschnitte werden den empirischen Forschungsstand aufarbeiten, die Methodik der Studie betrachten sowie die Ergebnisse und deren Implikationen diskutieren. Die Erläuterungen in den nachfolgenden Unterabschnitten stammen größtenteils aus dem bereits veröffentlichten Artikel „An Empirical Study on the Usage of Social Media in German B2C-Online Stores“ [90].

3.3.1 Empirischer Forschungsstand

Es gab in der Vergangenheit bereits mehrere Untersuchungen die sich mit dem Einfluss von Social Media auf den E-Commerce beschäftigt haben. Einige davon sollen nachfolgend kurz beschrieben und von der Untersuchung dieser Arbeit abgegrenzt werden.

Ein Beitrag von Stormer und Frauchinger aus dem Jahr 2008 hat gezeigt, dass einige Onlineshops eine geringe soziale Vernetzung der Konsumenten aufweisen [186, S. 69]. Die Hälfte der insgesamt 12 untersuchten Onlineshops wurde mit einer geringen sozialen Interaktion der Konsumenten eingestuft. Lediglich der untersuchte Onlineshop Etsy¹ wurde mit einer hohen sozialen Vernetzung ausgezeichnet. Aufgrund der geringen Anzahl an untersuchten Onlineshops und der relativ groben Einteilung in geringe, mittlere oder hohe soziale Vernetzung, ist die Aussagekraft dieses Beitrags jedoch als eher gering einzustufen.

Bereits im Jahr 2009 untersuchten Leitner und Grechenig neue Formen der Interaktion und Kollaboration im E-Commerce. Ihr Ansatzpunkt waren dabei 100 internationale Social Onlineshopping Webseiten. Die Ergebnisse der Untersuchung sind teilweise überraschend, da Diskussionsmöglichkeiten in den untersuchten Shops nur schlecht vertreten waren, z.B. Foren nur zu 31 % und Chats sogar nur zu 2 %. Es ist zu erwähnen, dass die Onlineshopping Webseiten keine eigenen Produkte anbieten, sondern nur eine Plattform zur sozialen Vernetzung und Weiterleitung an die konkreten Onlineshops bereitstellen. Ferner wurde neben dem B2C-Segment ebenfalls das C2C-Segment betrachtet. Diese beiden Punkte unterscheiden die Untersuchung von der nachfolgenden Untersuchung dieser Arbeit. Der Untersuchungsgegenstand dieser Arbeit sind klassische B2C-Onlineshops, welche Güter eigenständig vertreiben.

Eine weitere ähnliche Studie wurde von Lackermair und Reuder durchgeführt, wobei sich deren Analyse von 100 Onlineshops auf die Kategorisierung und die Merkmale von Web 2.0-Anwendungen fokussiert hat [111]. Die zu untersuchenden Merkmale waren nicht explizit auf eine soziale Interaktion zwischen den Konsumenten, d.h. auf Social Media Features, ausgerichtet, sondern umfassten beispielsweise auch das Vorhandensein von Rich Media Inhalten oder Podcasts. Eine weitere Abgrenzung zu dieser Arbeit ist, dass in der Studie von Lackermair und Reuder feste Kategorien für die Erfassung der Social Media Features definiert wurden, z.B. Kommentarfunktionalitäten oder Tagging. Die Studie in dieser Arbeit beschränkt sich nicht auf fest vorgegebene Kategorien, sondern versucht exploratorisch alle bereitgestellten Social Media Features eines Onlineshops zu identifizieren.

Abschließend ist die Studie von Huang et al. zu erwähnen, welche 20 Onlineshops aus den USA und Kanada hinsichtlich der Verwendung von Social Media Features untersucht hat [81]. Das Konzept der Studie ist ähnlich zu der Studie in dieser Arbeit, da es aus Anwendersicht die Social Media Features von B2C-Onlineshops untersucht. Allerdings weicht deren Kategorisierung der gefundenen Social Media Features von dieser Studie ab. Huang et al. definieren insgesamt fünf Kategorien, wobei es sich bei einer davon um Social Media Marketing handelt. Eine solche Kategorie wäre eher in einer Verkäufer-orientierten Studie, als in einer Anwender-orientierten Studie zu erwarten. Die Klassifizierung der Studie in dieser Arbeit basiert auf den in der Literatur anerkannten Social Media Bausteinen nach Kietzmann et al. [97], welche bereits in Abschnitt 3.1.2 näher erläutert wurden.

¹<http://www.etsy.com> (besucht am 01.04.2014)

3.3.2 Methodisches Vorgehen

Stichprobe

Aufgrund der Vielzahl an B2C-Onlineshops war für die Studie eine Vollerhebung weder möglich noch sinnvoll. Daher sollte sich die Studie auf eine Stichprobe der Grundgesamtheit beziehen (Teilerhebung). Die beschriebene Grundgesamtheit liegt allerdings nicht vollständig erfasst vor, d.h., es existiert kein Verzeichnis, aus dem alle deutschen B2C-Onlineshops ermittelt werden könnten. Daher wurden für die Studie systematisch die 100 umsatzstärksten B2C-Onlineshops aus Deutschland für die Jahre 2010 und 2011 herangezogen. Die Datenbasis hierfür lieferten die Webseiten von iBusiness^{1,2}. Das Kriterium des Umsatzes wurde für die Stichprobe gewählt, da hierdurch von einer gewissen Reife des Onlineshops ausgegangen werden kann. Ferner können diese Onlineshops als relevant betrachtet werden, da anhand des hohen Umsatzes davon ausgegangen wird, dass diese Onlineshops eine Vielzahl von Konsumenten betreffen.

Für die Stichprobe wurden die umsatzstärksten Onlineshops der Jahre 2010 und 2011 verknüpft. Viele der Onlineshops von 2011 waren bereits 2010 in den Top 100 präsent, so dass sich nach Vereinigung der beiden Mengen eine Gesamtanzahl von 117 verschiedenen Onlineshops ergab. Eine Überprüfung vor Beginn der Studie hat gezeigt, dass die beiden Onlineshops von Neckermann³ und Schlecker⁴ zum Zeitpunkt der Studie nicht mehr existent waren, wodurch die endgültige Anzahl der untersuchten Onlineshops 115 betrug. Eine Liste dieser 115 Onlineshops ist im Anhang in Tabelle A.1 (siehe Seite 221) zu finden.

Da das Sortiment der zu untersuchenden Onlineshops sehr heterogen ist, wurde eine Einteilung in 10 Kategorien vorgenommen. Sortimente die weniger als fünfmal in der Stichprobe enthalten waren, wurden der Kategorie „Sonstiges“ zugewiesen. Ein Überblick zu den Kategorien und deren Vorkommnissen ist Abbildung 3.2 zu entnehmen.

Datenerhebung

Zur Erhebung der Daten wurde eine Beobachtungsstudie basierend auf der im vorigen Abschnitt genannten Stichprobe durchgeführt. Der Beobachtungszeitraum startete am 15.10.2012 und endete am 06.11.2012. Als Webbrowser wurde Mozilla Firefox⁵ in der Version 16.0.2 verwendet. Die Webseiten der Onlineshops wurden nach einem standardisierten Verfahren untersucht, welches für jeden Onlineshop wie folgt aussah:

1. Webseite des Onlineshops öffnen und Benutzerkonto anlegen (falls möglich)

¹http://www.ibusiness.de/lib/files/iBusiness_Poster_Onlineshops_2011.pdf (besucht am 17.09.2014)

²http://www.ibusiness.de/lib/files/iBusiness-Poster-Shopping-Portale_und_Online-Shops_2012.pdf.html (besucht am 17.09.2014)

³<http://www.neckermann.de> (besucht am 01.04.2014)

⁴<http://www.schlecker.de> (besucht am 01.04.2014)

⁵<http://mozilla.org/firefox> (besucht am 01.04.2014)

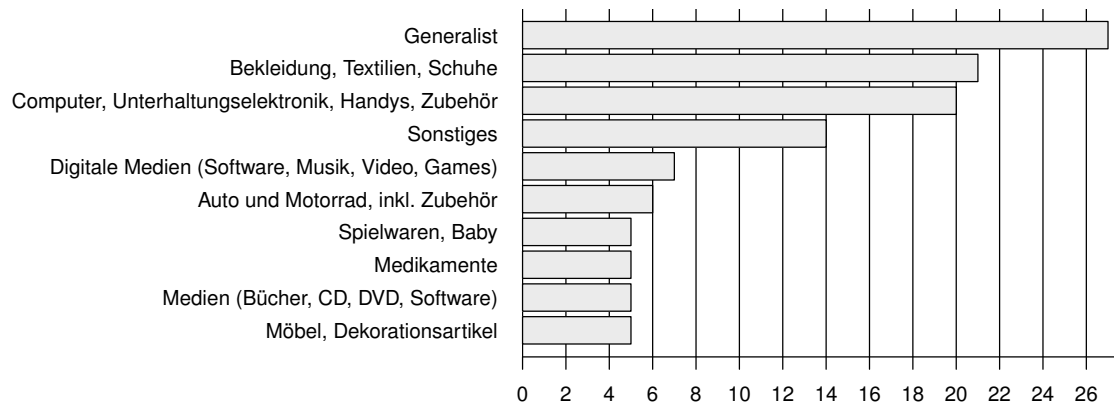


Abbildung 3.2: Sortimentsverteilung der untersuchten Onlineshops

2. Startseite des Onlineshops sowie die Seite des Benutzerprofils hinsichtlich Social Media Features analysieren
3. Eine zufällig ausgewählte Produktseite des Onlineshops öffnen und die dort vorhandenen Social Media Features identifizieren

Schritt 1: Beim ersten Besuch des Onlineshops wurde nach einer Möglichkeit gesucht, um ein Benutzerkonto einzurichten. Dies war notwendig, damit alle Social Media Features, also auch diejenigen die nur für registrierte Benutzer zugänglich sind, erfasst werden konnten. Bei fünf Onlineshops¹ war eine Registrierung nicht möglich, da hier die Angabe von sensiblen Informationen (z.B. Kreditkartennummer) oder die tatsächliche Durchführung eines Einkaufsvorgangs notwendig gewesen wäre. Daher besteht die Möglichkeit, dass für diese Onlineshops nicht alle Social Media Features identifiziert wurden.

Schritt 2: Nach der Erzeugung eines Benutzerkontos wurde die Startseite des Onlineshops nach Indikatoren für Social Media Features durchsucht, z.B. Links zu Blogs oder Diskussionsforen. Diese Features wurden dann analysiert und im Beobachtungsprotokoll vermerkt. Danach wurde die Profilseite des Benutzerkontos in ähnlicher Form analysiert.

Schritt 3: Abschließend wurde eine zufällig ausgewählte Produktseite des Onlineshops hinsichtlich Social Media Features, z.B. auf Vorkommnisse von Kundenrezensionen oder Produktbewertungen, analysiert.

Bei der Beobachtung der Onlineshops wurden alle Social Media Features notiert und später als dichotome Attribute modelliert, d.h., ein Feature kommt in einem Onlineshop vor oder nicht. Die Summe aller verschiedenen Features wird als Menge F gekennzeichnet. Später wurde jedes Social Media Feature einem der sieben Bausteine des Social Media

¹HSE24, Esprit, Baur, Klingel und ZooPlus

Tabelle 3.1: Lageparameter für die Verteilung der Social Media Features

Minimum	1. Quartil	Median	Mittelwert	3. Quartil	Maximum
0	3	5	5,05	7	14

zugeordnet (siehe hierzu auch Abschnitt 3.1.2). Ferner wurde eine weitere Kategorisierung vorgenommen, indem jedes Social Media Feature einem allgemeinen Bereich (F_A) oder Produkt-spezifischen Bereich (F_P) des Onlineshops zugewiesen wurde. Die Menge F_P enthält daher nur Features, die auf Produktseiten gefunden wurden.

Hinsichtlich des verwendeten Vorgehens sind abschließend noch zwei Einschränkungen zu berücksichtigen. Da im Rahmen der Beobachtung keine echten Käufe getätigt wurden, konnten auch nur Social Media Features erfasst werden, welche vor einem Kauf zur Verfügung standen. Alle nachträglichen Features, z.B. das Teilen eines Einkaufs über soziale Netzwerke, konnten daher nicht erfasst werden. Ferner wurden nur Social Media Features erfasst, welche auch tatsächlich eine Benutzerinteraktion erforderten. Passive Empfehlungsansätze, die beispielsweise auf Collaborative Filtering beruhen, wurden daher nicht erfasst.

3.3.3 Ergebnisse

Für die 115 untersuchten Onlineshops konnten insgesamt 25 unterschiedliche Social Media Features gefunden werden ($F = 25$). Darauf entfallen 11 Features die im allgemeinen Bereich der Onlineshops gefunden wurden ($F_A = 11$) sowie 14 Features die im Produkt-spezifischen Bereich gefunden wurden ($F_P = 14$).

Insgesamt wurden auf allen Onlineshops 581 Verwendungen von Social Media Features festgestellt, was zu einem Mittelwert von 5,05 eingesetzten Features pro Onlineshop führt. Von den 581 gefundenen Features wurden 127 im allgemeinen Bereich und 454 im Produkt-spezifischen Bereich der Onlineshops gefunden. Es gab sechs Onlineshops die keine Social Media Features bereitgestellt haben und einen Anbieter mit einem Maximum von 14 bereitgestellten Features, wobei dieser Wert als Ausreißer bezeichnet werden kann (vgl. Boxplot aus Abbildung 3.3). Neben dem Boxplot enthält Abbildung 3.3 ebenfalls ein Histogramm, das die Verteilung der Social Media Features grafisch darstellt. Die Lageparameter zur Verteilung der Werte sind Tabelle 3.1 zu entnehmen.

Die gefundenen Social Media Features aus dem allgemeinen Bereich sind in Tabelle 3.2 aufgelistet. Das am häufigsten verwendete Feature in diesem Bereich ist ein *Weblog* (kurz *Blog*). Fast jeder dritte Onlineshop verfügt über einen eigenen Blog. Circa 26 % der Onlineshops erlauben die Einrichtung einer Produkt-Wunschliste, zu der Konsumenten Produkte hinzufügen können. In einem Onlineshop war es auch möglich, die Wunschlisten von anderen Kunden zu bewerten. Diskussionsplattformen, wie z.B. Foren, waren in rund 18 % der Onlineshops anzutreffen. In 16 % der Fälle wurde die Erzeugung eines eigenen, öffentlichen Benutzerprofils ermöglicht. Etwa 9 % der Onlineshops stellten ein System bereit, das die Bewertung von anderen Benutzern (z.B. hinsichtlich ihrer Rezensionen) erlaubt. Etwas seltener genutzte Interaktionsfeatures umfassten die Verwaltung von

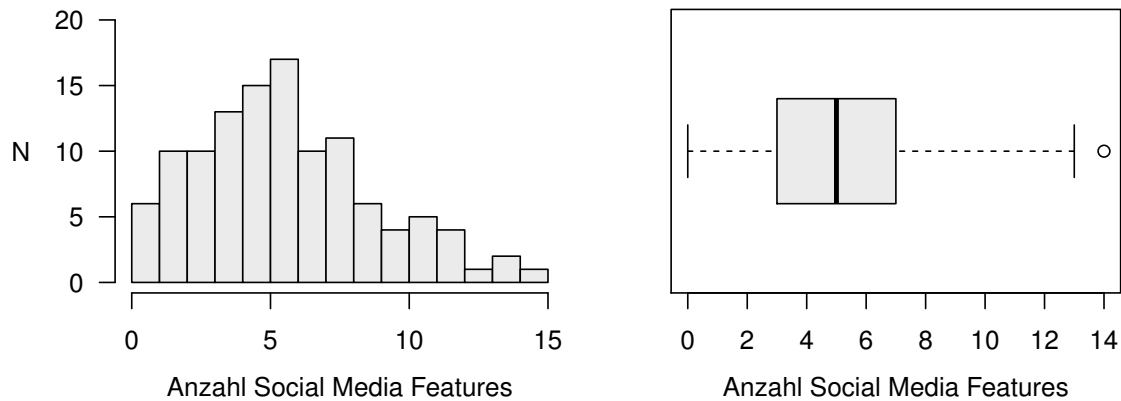


Abbildung 3.3: Histogramm und Boxplot für die Verteilung der Social Media Features in den untersuchten Onlineshops

Tabelle 3.2: Social Media Features aus dem allgemeinen Bereich (F_A) und deren Verwendungshäufigkeit

Feature	Anteil
Blog	33,04 %
Wunschliste mit anderen teilen	26,09 %
Diskussionsplattformen (z.B. Foren)	18,26 %
Öffentliche Kundenprofile	15,65 %
Personen-bezogene Bewertungen (z.B. Top Rezensent)	8,70 %
Nachrichten an andere Nutzer senden	2,61 %
Kontakte innerhalb des Onlineshops verwalten	1,74 %
Nutzer-spezifische Blogs	1,74 %
Online-Status anderer Nutzer bestimmen	0,87 %
Ereignisse/Kalender verwalten	0,87 %
Wunschliste von anderen Nutzern bewerten	0,87 %

Kontakten (ca. 2 %) und das Senden von Nachrichten (ca. 3 %). Die Erkennung des Online-Status von anderen Mitgliedern sowie die Verwaltung von Ereignissen innerhalb eines Kalenders wurden von jeweils einem Onlineshop bereitgestellt.

Im Produkt-spezifischen Bereich, d.h. auf den Produktseiten der Onlineshops, wurden 14 verschiedene Features gefunden, welche in Tabelle 3.3 dargestellt sind. Die zwei am Häufigsten genutzten Features sind hier Produktbewertungen (ca. 69 %) und Kundenrezensionen (ca. 66 %), welche in Abschnitt 3.4 noch näher betrachtet und differenziert werden. In ca. 96 % der Fälle wurden Produktbewertungen und Kundenrezensionen gemeinsam angeboten, d.h., nur etwa 4 % der Onlineshops haben Produktbewertungen und Kundenrezensionen getrennt voneinander angeboten.

40 % der Onlineshops boten die Möglichkeit zur Bewertung der Nützlichkeit von Kundenrezensionen. Ein Onlineshop erlaubte auch die Möglichkeit einem Rezensenten zu

Tabelle 3.3: Social Media Features aus dem Produkt-spezifischen Bereich (F_P) und deren Verwendungshäufigkeit

Feature	Anteil
Produktbewertungen	68,70 %
Kundenrezensionen	66,09 %
Produktempfehlungen per E-Mail versenden	59,13 %
Produktdetails über soziale Netzwerke teilen (z.B. Facebook)	53,91 %
Produkte bewerten über soziale Netzwerke (z.B. Facebook)	48,70 %
Rezensionen als hilfreich zu kennzeichnen	40,00 %
Detaillierte Produktbewertungen (zusätzlich zur Gesamtbewertung)	18,26 %
Vor- und Nachteile zu Rezensionen hinzufügen	10,43 %
Upload von Benutzer-generierten Inhalten (z.B. Produktbilder)	10,43 %
Kommentare zu Kundenrezensionen	8,70 %
Persönliche Daten zu Rezensionen anfügen	8,70 %
Benutzer-definierte Schlagwörter (Tags)	3,48 %
Eigene „Gefällt mir“-Funktion für Produkte	0,87 %
Rezensent für Rezension zu danken	0,87 %

danken, wodurch die Reputation der Person verbessert wurde. Die Kommentierung von Kundenrezensionen war in ca. 9 % der Fälle möglich. Eine detaillierte Produktbewertung (zusätzlich zur Gesamtbewertung) wurde in ca. 18 % der Fälle ermöglicht. Dabei konnten mehrere vordefinierte Produktmerkmale einzeln bewertet werden. Im Vergleich dazu erlaubten ca. 10 % die Angabe von Vor- und Nachteilen als Freitext. Zur besseren Interpretation von Kundenrezensionen erlaubten ca. 9 % der Onlineshops die Angabe von persönlichen Daten des Rezensenten, z.B. Alter oder Hauttyp für Produkte aus dem Bereich Kosmetik.

Fast 60 % der Onlineshops unterstützen eine Produktempfehlung via E-Mail, bei der ein Kunde den Uniform Resource Locator (URL) einer Produktseite per E-Mail an einen Bekannten weiterleiten kann. Das Teilen von Produktdetails über soziale Netzwerke (z.B. Facebook oder Twitter) war in etwa jedem zweiten Onlineshop möglich. Auch die Einbindung von Empfehlungsbuttons aus sozialen Netzwerken (z.B. der „Like“-Button) war in fast der Hälfte aller Fälle möglich. Ein Onlineshop stellte auch einen eigenen „Gefällt mir“-Button zur Verfügung. Die Erstellung von Benutzer-definierten Schlagwörtern (Tags) war nur in 3 % der Fälle möglich. Etwa 10 % der Onlineshops erlaubten das Hochladen von Benutzerinhalten, z.B. Produktfotos.

Da es sich bei den Produktbewertungen um das häufigste Feature handelt, welches insgesamt von 79 der 115 untersuchten Onlineshops eingesetzt wird, wurde dieses Feature näher untersucht. Es hat sich gezeigt, dass 76 der 79 Onlineshops eine Bewertungsskala von 1–5 anbieten. Die restlichen 3 Onlineshops verwendeten eine 10-stufige Bewertungsskala. Von den 76 Onlineshops mit der 5-stufigen Bewertungsskala nutzen 72 (91,14 %) als Symbol einen Stern, also die sogenannte 5-Sterne Bewertungsskala.

Tabelle 3.4: Durchschnittliche Anzahl an Social Media Features unterteilt nach Sortimenten

Sortiment	Features	Onlineshops
Auto und Motorrad, inkl. Zubehör	1,83	6
Bekleidung, Textilien, Schuhe	4,86	21
Computer, Unterhaltungselektronik, Handys, Zubehör	4,95	20
Digitale Medien (Software, Musik, Video, Games)	2,86	7
Generalist	5,74	27
Medien (Bücher, CD, DVD, Software)	6,80	5
Medikamente	3,40	5
Möbel, Dekorationsartikel	3,60	5
Sonstiges	6,71	14
Spielwaren, Baby	6,20	5

3.3.4 Diskussion

Die Ergebnisse lassen interessante Rückschlüsse zu. So befinden sich beispielsweise 78 % der Social Media Features auf den Produktseiten der Onlineshops, was vermuten lässt, dass Onlineshops in diesem Bereich eine erhöhte Interaktion erwarten. Der Großteil der hier genutzten Features (55 %) betrifft Produktbewertungen, Kundenrezensionen und alle damit verbundenen Features, z.B. die Kommentierung von Rezensionen. Dies zeigt, dass diese beiden Features in den untersuchten Onlineshops am häufigsten vertreten waren. Es lässt sich auch folgern, dass Anbieter Produktbewertungen und Kundenrezensionen als komplementär betrachten, denn 94 % der Onlineshops bieten sowohl Produktbewertungen als auch Kundenrezensionen an. Ein eher überraschendes Ergebnis war, dass die Nutzung von Schlagwörtern (Tags) in nur etwa 3 % aller Fälle bereitgestellt wurde, obwohl dies eine flexible Kategorisierung durch Anwender ermöglichen würde.

Wie bereits in Abschnitt 3.3.1 erwähnt wurde, ist die Studie von Huang et al. ähnlich zu dieser Studie. Obwohl die Vergleichbarkeit mit dieser Studie aufgrund der abweichenden methodischen Ansätze schwierig ist, lassen sich doch ähnliche Ergebnisse festhalten. Die am häufigsten verwendeten Features in deren Studie waren ebenfalls Produktbewertungen, Kundenrezensionen und die integrierten Features aus sozialen Netzwerken, wie z.B. „Teilen“- oder „Gefällt mir“-Buttons [81].

Für eine bessere Kategorisierung der gefundenen Social Media Features wurden diese den sieben Social Media Bausteinen aus Abbildung 3.1 zugeordnet (siehe Tabelle 3.5). Wie der Tabelle zu entnehmen ist, wurden insgesamt nur sechs der sieben Bausteine identifiziert, wobei der Baustein „Gruppen“ in keinem der untersuchten Onlineshops genutzt wurde. Der Baustein „Austausch“ lässt sich noch weiter in eine interne und externe Komponente unterteilen. Beim internen Austausch werden Inhalte mit anderen Benutzern des Onlineshops geteilt, während der externe Austausch meist über soziale Netzwerke (wie z.B. Facebook) stattfindet. Vergleicht man die Social Media Features des Bausteins „Austausch“ mit Tabelle 3.3 so zeigt sich, dass ein externer Austausch von Inhalten häufiger

Tabelle 3.5: Zuteilung der gefundenen Social Media Features zu den Bausteinen des Social Media

Baustein	Social Media Feature
Konversation	Blog Diskussionsplattformen (z.B. Foren) Kommentare zu Kundenrezensionen Nachrichten an andere Nutzer senden Nutzer-spezifische Blogs
Identität	Öffentliche Kundenprofile
Präsenz	Online-Status anderer Nutzer bestimmen
Beziehungen	Kontakte innerhalb des Onlineshops verwalten
Reputation	Vor- und Nachteile zu Rezensionen hinzufügen Eigene „Gefällt mir“-Funktion für Produkte Produktbewertungen Kundenrezensionen Detaillierte Produktbewertungen (zusätzlich zur Gesamtbewertung) Produkte bewerten über soziale Netzwerke (z.B. Facebook) Personen-bezogene Bewertungen (z.B. Top Rezensent) Rezensionen als hilfreich zu kennzeichnen Wunschliste von anderen Nutzern bewerten Rezensent für Rezension zu danken
Austausch	Persönliche Daten zu Rezensionen anfügen Ereignisse/Kalender verwalten Produktempfehlungen per E-Mail versenden Wunschliste mit anderen teilen Produktdetails über soziale Netzwerke teilen (z.B. Facebook) Upload von Benutzer-generierten Inhalten (z.B. Produktbilder) Benutzer-definierte Schlagwörter (Tags)

genutzt wird als ein interner Austausch. Dies führt zur Vermutung, dass Onlineshops mehr daran interessiert sind, neue Kunden über externe Empfehlungen zu akquirieren, als interne Interaktionsfunktionalitäten für bestehende Kunden bereitzustellen.

Nach der Zuordnung der Social Media Features kann die Nutzung der einzelnen Bausteine analysiert werden. Abbildung 3.4 zeigt wie viele Prozent der 115 Onlineshops die einzelnen Social Media Bausteine genutzt haben. Der am häufigsten genutzte Baustein ist „Austausch“. Die Grafik aus Abbildung 3.4 zeigt, dass 93 der 115 Onlineshops (ca. 81 %) mindestens ein Social Media Feature aus dem Baustein „Austausch“ verwendet haben. Vergleichsweise oft wurde der Baustein „Reputation“ verwendet. Die Bausteine „Identität“, „Präsenz“ und „Beziehungen“ spielen bei den untersuchten Onlineshops eine untergeordnete Rolle. Der Baustein „Gruppen“ wurde, wie bereits erwähnt, von keinem der Onlineshops verwendet.

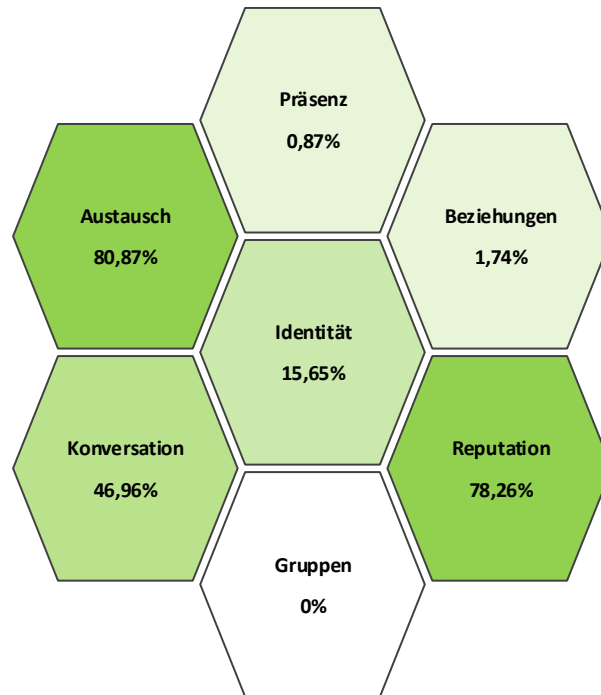


Abbildung 3.4: Verteilung der gefundenen Social Media Features auf die sieben Bausteine des Social Media

Abschließend müssen für die durchgeführte Studie noch Einschränkungen gemacht werden. Zum einen wurden in der Studie nur die umsatzstärksten B2C-Onlineshops untersucht. Die Social Media Charakteristiken für kleinere bzw. umsatzschwächere Onlineshops könnten von den Ergebnissen der Studie abweichen. Ferner sagen die Ergebnisse der Studie nichts über die tatsächliche Nutzung der Social Media Features aus. Der Rahmen der Studie adressierte letztlich nur die Bereitstellung dieser Dienste aus Sicht der Onlineshops und nicht die Nutzung durch die Besucher der Onlineshops.

Basierend auf der empirischen Studie und dem aktuellen Stand der Forschung ist der in Kapitel 4 beschriebene Social Media Dienst entstanden. Der Dienst basiert auf einigen in der Studie analysierten Social Media Features, nämlich der Produktbewertung, Kundenrezension und dem kollaborativen Tagging. Aufgrund der Bedeutung dieser Komponenten werden diese in den nachfolgenden Abschnitten noch detaillierter betrachtet.

3.4 Online-Bewertungen

Wie eine Umfrage im Auftrag von BITKOM aus dem Jahr 2013 ergeben hat, nimmt die Bedeutung von Online-Bewertungen im Onlinehandel stetig zu [17]. Laut dieser Umfrage bewertet bereits jeder zweite Nutzer Produkte oder Dienstleistungen im Internet. Im weiteren Sinne umfasst der Begriff „Online-Bewertung“ typischerweise zwei Arten von Bewertungen. Zum einen existiert ein quantifizierbares Verfahren bei dem Bewertungen

zahlen- oder wertmäßig erfasst werden können, z.B. mithilfe einer Bewertungsskala. Solche Bewertungen werden in der Literatur auch als *Produktbewertungen* (engl. *product ratings*) oder nur als *Ratings* bezeichnet [66]. Zum anderen können Bewertungen auch textuell, mithilfe der sogenannten *Produktrezensionen* (engl. *product reviews*), vorgenommen werden. Diese beiden Bewertungsansätze werden in den nächsten beiden Abschnitten erläutert und voneinander abgegrenzt.

3.4.1 Produktbewertungen

Die zahlen- oder wertmäßige Erfassung einer Bewertung ist nicht nur im Onlinehandel verbreitet. Ein einfaches Beispiel hierfür ist die Bewertung von Inhalten auf Facebook anhand des „Gefällt mir“-Buttons. Jeder Klick auf diesen Button erhöht die Reputation des zu bewertenden Inhalts, d.h., die Reputation eines Inhaltes ist bestimmbar durch die Anzahl der Klicks auf diesen Button. Eine andere Variante wird auf dem Videodienst Youtube¹ verwendet. Hier können Nutzer einen Inhalt entweder als positiv oder als negativ bewerten. Anhand der Gegenüberstellung der Bewertungen lässt sich dann die Reputation eines Videos bestimmen.

Diese relativ grobe Aufteilung der Bewertungen funktioniert zwar für einige Inhalte, ist aber für Produktbewertungen im Onlinehandel eher ungeeignet. Hier werden detailliertere Bewertungsansätze benötigt, um die Entscheidungsfindung von Konsumenten zu erleichtern. Als Bewertungsansatz für Produktbewertungen hat sich im Onlinehandel eine *Bewertungsskala* durchgesetzt. Bei diesem Bewertungsansatz wird dem Nutzer ein Wertebereich präsentiert, wobei der höchste Wert typischerweise der bestmöglichen Bewertung entspricht. Die in dieser Arbeit durchgeführte Studie (siehe Abschnitt 3.3) hat gezeigt, dass verschiedenste Ausprägungen für Bewertungsskalen existieren und dass die meisten der untersuchten Onlineshops als Bewertungsskala eine sogenannte *5-Sterne Bewertungsskala* verwenden. Bei dieser Skala kann der Nutzer einen Inhalt mit ein bis fünf Sternen bewerten, wobei fünf Sterne für die bestmögliche Bewertung stehen.

Bei der Bewertungsskala handelt es sich um eine Intervallskala, d.h., die Abstände zwischen den Werten werden als gleich groß erachtet. Es handelt sich jedoch nicht um eine Verhältnisskala, denn ein Artikel mit einer Bewertung von vier Sternen ist nicht zwingend doppelt so gut wie ein Artikel mit einer Bewertung von zwei Sternen. Basierend auf den Eigenschaften einer Intervallskala ist auch die Bildung eines arithmetischen Mittels zur Aggregation von mehreren Bewertungen erlaubt [21, S. 20]. Dies ist besonders hilfreich um Konsumenten einen ersten Überblick über die durchschnittliche Bewertung eines Produkts oder einer Dienstleistung zu geben.

Neben der Aggregation können Produktbewertungen auch zur Sortierung und Filterung verwendet werden, beispielsweise zur Reduzierung des Consideration Sets [77]. Dies zeigt, dass Produktbewertungen von Konsumenten typischerweise in einer früheren Phase des Entscheidungsprozesses eingesetzt werden. Eine Filterung ist jedoch auch in späteren

¹<http://www.youtube.com> (besucht am 06.04.2014)



Abbildung 3.5: Detaillierte Bewertung von vordefinierten Produktaspekten (Quelle: <http://www.conrad.de>, besucht am 07.04.2014)

Phasen sinnvoll, z.B. um nur schlechte Kundenbewertungen (weniger als drei von fünf Sternen) anzuzeigen.

Produktbewertungen betreffen meistens das Produkt oder die Dienstleistung als Ganzes und werden daher oft in Onlineshops als Gesamtbewertung ausgezeichnet. Es existieren jedoch auch Ansätze um zusätzlich zur Gesamtbewertung auch noch einzelne Aspekte zu bewerten. In der durchgeführten Studie haben allerdings nur etwa 18 % der Onlineshops eine solche Bewertungsmöglichkeit angeboten. Ein Beispiel für eine detailliertere Bewertung von Produktaspekten ist in Abbildung 3.5 dargestellt. Die Abbildung zeigt, dass neben der Gesamtbewertung auch Merkmale wie Benutzerfreundlichkeit oder Design auf einer 5-Sterne Skala bewertet werden können.

3.4.2 Produktrezensionen

Definition und Abgrenzung

Der Begriff *Produktrezension* (manchmal auch Kundenrezension oder Online-Rezension genannt [211]) wird häufig mit einer Produktbewertung gleichgesetzt. Dies ist jedoch nicht korrekt, denn im Gegensatz zur Produktbewertung ist eine Produktrezension nicht quantifizierbar, sondern liegt in einer textuellen Form vor. Unterscheidbar sind hier Produktrezensionen die von Produktexperten verfasst werden und solche die von tatsächlichen Endkunden verfasst werden. Zur ersten Kategorie zählt beispielsweise die Webseite CNET¹, welche anhand von Produkttests detaillierte Produktrezensionen verfasst. Für diese Forschungsarbeit hingegen sind nur die von Endnutzern verfassten Produktrezensionen interessant. Daher beziehen sich alle nachfolgenden Verwendungen des Begriffes auf diese Zielgruppe.

Die von Endnutzern erstellen Produktrezensionen können ferner hinsichtlich ihres Verwendungsorts abgegrenzt werden. Hier sind typischerweise zwei Verwendungen zu unterscheiden, nämlich Produktrezensionen auf sogenannten *Bewertungsportalen* und Produktrezensionen in elektronischen Marktplätzen. Bewertungsportale (wie z.B. Ciao²) stellen einen dedizierten Platz zur Bereitstellung von Kunden-basierten Produktrezensio-

¹<http://www.cnet.de> (besucht am 11.04.2014)

²<http://www.ciao.de> (besucht am 11.04.2014)

nen dar. Ein Verkauf von Produkten oder Dienstleistungen findet auf Bewertungsportalen meist nicht statt. Ein weiterer Einsatz von Produktrezensionen findet in elektronischen Marktplätzen, wie z.B. in Onlineshops oder Internetauktionshäusern, statt. Diese Forschungsarbeit fokussiert sich auf die Verwendung von Produktrezensionen in Onlineshops.

In der Forschungsliteratur werden Produktrezensionen oft unter dem Themenkomplex *electronic Word-Of-Mouth (eWOM)* zusammengefasst [29, 211]. Der Begriff eWOM steht für eine elektronische Mundpropaganda, welche in textueller Form vorliegt, von einer Person für andere Personen verfasst ist und die Leistungsfähigkeit von Produkten oder Dienstleistungen aus Sicht des Verfassers beschreibt [211]. Der Bereich eWOM geht also über Produktrezensionen hinaus und umfasst auch Empfehlungen die in Blogs oder in sozialen Netzwerken, wie beispielsweise Facebook, zum Ausdruck gebracht werden. Im Rahmen dieser Forschungsarbeit sind allerdings nur Produktrezensionen innerhalb von Onlineshops relevant.

Wie die Studienergebnisse dieser Arbeit gezeigt haben (siehe Abschnitt 3.3), gibt es nur wenige Onlineshops die Produktrezensionen einzeln, d.h. ohne dazugehörige Bewertungskomponente, verwenden. Häufiger treten Produktrezensionen gepaart mit Produktbewertungen auf, wodurch der Konsument anhand der Produktbewertung einen ersten Eindruck von der Rezension erhält, z.B. ob die Meinung des Rezensenten eher positiv oder negativ ausfällt.

Einfluss auf Konsumenten

Produktrezensionen spielen für viele Konsumenten eine wichtige Rolle. Eine Befragung von 104 deutschen Onlinekunden hat ergeben, dass Produktrezensionen von etwa 74 % der befragten Personen als wichtig bzw. sehr wichtig angesehen werden [110]. Die Studie von Park et al. hat gezeigt, dass die Kaufabsichten eines Konsumenten durch die Qualität von Produktrezensionen sowie durch die Anzahl an verfügbaren Produktrezensionen positiv beeinflusst werden [151]. Dies zeigt, dass Produktrezensionen in die Entscheidungsfindung von Konsumenten einbezogen werden. Nach Park und Lee spielen Produktrezensionen hierfür zwei Rollen. Zum einen treten sie als *Informant* auf, d.h., sie liefern den Lesern der Rezension neue Informationen und zum anderen nehmen sie eine *empfehlende Funktion* ein, indem sie durch positive oder negative Aussagen die Leser der Rezension beeinflussen [150].

Der Einfluss von Produktrezensionen kann sich jedoch auch negativ auf die Entscheidungsfindung auswirken, wenn durch eine Vielzahl an Rezensionen eine Informationsüberflutung der Konsumenten entsteht. Dies hat sich auch in der Studie von Park und Lee gezeigt [150]. Es existieren jedoch auch Ansätze, die auf der automatischen Analyse von Produktrezensionen basieren und so den kognitiven Aufwand von Anwendern reduzieren sollen. Einige dieser Ansätze werden in Abschnitt 3.4.2 näher betrachtet.

Für den Einfluss von Rezensionen auf Konsumenten ist ebenfalls die Beschaffenheit der Produkte bzw. Dienstleistungen ausschlaggebend. Ricci und Wietsma haben in einer Studie aus dem Bereich des Tourismus herausgefunden, dass die Rolle der Rezensionen von der Beschaffenheit des Produkts abhängig ist [165, S. 299]. Sie beschreiben, dass für erfahrungsbasierte Produkte (wie z.B. Freizeitaktivitäten), welche schwer anhand von Pro-

duktmerkmalen modelliert werden können, die Rezensionen schon in einer früheren Phase des Entscheidungsprozesses eine größere Rolle spielen. Bei Produkten mit einem reichhaltigen Satz an Produktmerkmalen beeinflussen Rezensionen die Entscheidungsfindung der Konsumenten eher zu einem späteren Zeitpunkt.

Ferner haben sich einige empirische Untersuchungen mit der Frage beschäftigt, ob Produktrezensionen die Anzahl der Verkäufe beeinflussen. Hier sind sich die Autoren der bestehenden Arbeiten allerdings nicht einig, was an der Schwierigkeit in der Messbarkeit dieses Konstrukts liegen könnte. Die Studie von Chevalier und Mayzlin deutet auf einen positiven Einfluss von Produktrezensionen hinsichtlich der Verkaufsförderung hin [29], während Chen et al. sowie Duan et al. in ihren Studien keinen kausalen Zusammenhang entdeckt haben [27, 44].

Analyse von Produktrezensionen

Durch die steigende Beliebtheit von Online-Bewertungen wächst auch die Anzahl der Produktrezensionen. Dieser Anstieg wurde auch in der Studie von Chevalier und Mayzlin beobachtet, welche eine Stichprobe an Büchern auf Amazon analysiert hat. Innerhalb eines Jahres (Mai 2003 bis Mai 2004) hat sich die durchschnittliche Anzahl der Produktrezensionen pro Buch von 60,99 auf 68,31 gesteigert [29, S. 347]. Eine aktuellere Zahl bezüglich der durchschnittlichen Anzahl von Produktrezensionen im Tourismusbereich liefert die Touristikwebseite TripAdvisor, welche unter anderem Rezensionen für Hotels umfasst. Wie ein Pressebericht aus dem Jahr 2013 gezeigt hat, enthalten die 500 meist gesuchtesten Reiseziele durchschnittlich 139 Rezensionen [194]. Die genannten Zahlen machen deutlich, dass die Analyse dieser Vielzahl an Rezensionen für Konsumenten einen hohen Aufwand darstellt und eine komplette Analyse für alle infrage kommenden Produkte nicht sinnvoll wäre. Zur Verbesserung dieser Situation wurden verschiedene Ansätze entwickelt, wovon einige nachfolgend näher vorgestellt werden.

Ein Ansatz ist die Bereitstellung eines Mechanismus für Konsumenten, um Produktrezensionen als hilfreich oder nicht hilfreich zu bewerten. Über diese Möglichkeit können einem Konsumenten die am besten bewerteten Produktrezensionen zuerst präsentiert werden. Wie die Studie dieser Arbeit gezeigt hat, unterstützen 40 % der untersuchten Onlineshops diesen Ansatz (siehe Tabelle 3.3 auf Seite 39). Die Nützlichkeit einer Rezension ist allerdings immer subjektiv und basiert auf den Anforderungen der jeweiligen Konsumenten. So kann es vorkommen, dass Konsumenten mit anderen Anforderungen, eine als hilfreich bewertete Rezension möglicherweise nicht als hilfreich empfinden. Daher trägt dieser Ansatz nur eingeschränkt zur Verbesserung des oben genannten Problems bei.

Im Gegensatz zu quantifizierbaren Produktbewertungen enthalten Produktrezensionen detailliertere Informationen, z.B. zu Produktmerkmalen. Aufgrund ihrer textuellen Struktur sind sie jedoch nur schwer auswertbar und aggregierbar. Es existieren jedoch Forschungsansätze, um diese textuellen Informationen in eine quantifizierbare Form zu bringen und so unter anderem eine Aggregation zu ermöglichen. Die Analyse solcher Textinhalte ist dem Forschungsbereich *Text Mining* zuzuordnen. Gemäß Kao und Poteet umfasst Text Mining „das Auffinden und Extrahieren von interessanten, nicht-trivialen

Informationen aus unstrukturierten Texten“ [95, S. 1]. Eine spezielle Anwendung des Text Mining, welche besonders für den Rahmen dieser Arbeit interessant ist, betrifft die Extraktion von Meinungen aus Texten, z.B. aus Produktrezensionen. Dieser Forschungsbereich wird unter dem Schlagwort *Opinion Mining* zusammengefasst [147, S. 10]. Laut Binali et al. ist das Ziel des Opinion Mining die Extraktion von Meinungen aus verschiedenen Quellen (z.B. Produktrezensionen) und die Darstellung dieser Informationen in einer benutzerfreundlichen Weise [16, S. 1].

Anhand des Opinion Mining soll unter anderem bestimmt werden, ob es sich um eine eher positive oder negative Produktrezension handelt. Da Produktrezensionen aber ohnehin sehr oft gepaart mit Gesamtbewertungen auftreten, liefert eine solche gesamtheitliche Betrachtung der Produktrezension jedoch meist nur einen geringen Mehrwert. Interessanter ist die Extraktion von Meinungen in Hinblick auf die in der Rezension beschriebenen Produktmerkmale. Dies wird von Liu als *Feature-based Opinion Mining* bezeichnet [118, S. 417]. Hierzu werden für alle Sätze der Rezension die folgenden Schritte durchgeführt [118, S. 418]:

1. Ermitteln und Extrahieren der Produktmerkmale, zu denen der Rezensent eine Meinung abgegeben hat
2. Ermitteln, ob die beschriebenen Produktmerkmale positiv, negativ oder neutral bewertet wurden

Anhand des folgenden Satzes soll die Verwendung der oben genannten Schritte verdeutlicht werden: „Ich finde die *Klangqualität* des Kopfhörers *sehr gut*.“ Die hervorgehobenen Satzteile stellen jeweils das Produktmerkmal (*Klangqualität*) sowie dessen Bewertung (*sehr gut*) dar. Natürlich stellt das genannte Beispiel einen einfach formulierten Satz dar, welcher auch leicht automatisiert auswertbar ist. Von einem korrekten Satzbau kann jedoch in Produktrezensionen nicht immer ausgegangen werden. Schwieriger wird es beispielsweise, wenn Sätze Rechtschreibfehler enthalten oder der komplette Satzbau fehlerhaft ist. So wäre beispielsweise der Satz „Klingt *ganz gut*.“ problematischer für eine Opinion Mining Anwendung, da dieser Satz nicht explizit ein Produktmerkmal, sondern nur eine explizite Bewertung (*ganz gut*) enthält.

3.5 Tagging

3.5.1 Grundlagen

Mit der immer größer werdenden Datenmenge in aktuellen Informationssystemen steigt auch die Schwierigkeit Inhalte wieder auffindbar zu machen. Eine Möglichkeit bietet hier unter anderem eine Inhaltsklassifikation mittels Schlagwörtern. Dabei weist ein Anwender einer Ressource passende, frei wählbare Schlagwörter (engl. *keywords* oder *tags*) zu. Dieser Prozess wird auch als Verschlagwortung oder *Tagging* bezeichnet [57]. Wie aus der genannten Definition zu erkennen ist, umfasst das Tagging die Komponenten *Anwender*, *Tag* und *Ressource* [180].

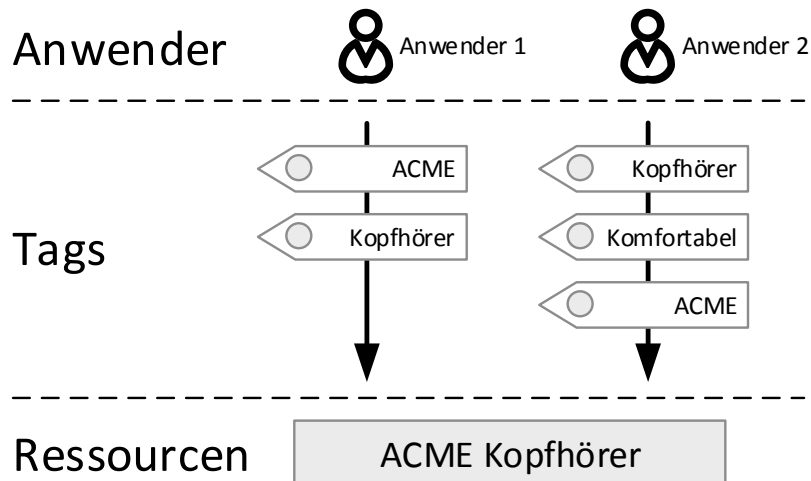


Abbildung 3.6: Beispielhaftes, vereinfachtes Modell eines Social Tagging Systems

Die Idee des Tagging ist es, starre Taxonomien durch flexible Metadaten zu kompensieren, da Ressourcen nicht immer nur in genau eine Kategorie passen [57]. Ein Beispiel sind die Produktkategorien im Onlinehandel. Hier wird einem Produkt typischerweise nur genau eine Produktkategorie zugewiesen, obwohl ein Produkt womöglich in mehrere Kategorien passen könnte. Ein anderes Beispiel wäre eine MP3-Musikdatei, welche in genau einem Ordner innerhalb des Dateisystems existiert. Über sogenannte ID3-Tags ist beispielsweise die Zuweisung von Metadaten (z.B. das Genre des Musiktitels) möglich, wodurch dem Anwender eine spätere Suche nach bestimmten Musikdateien erleichtert wird.

Je nach Aufbau eines Tagging Systems lassen sich in Anlehnung an Smith zwei Arten des Tagging identifizieren, nämlich das *einfache Tagging* und das *kollaborative Tagging* (auch *Social Tagging* genannt) [180]. Beim einfachen Tagging geschieht die Zuweisung von Tags zu einer Ressource nur durch den Besitzer der Ressource, d.h. nur durch einen Anwender. Ein Beispiel hierzu ist der Videodienst Youtube. Nur der Anwender der ein Video eingestellt hat kann auch dessen Tags vergeben. Andere Nutzer des Youtube-Dienstes können dem Video keine Tags zuweisen. Die zugewiesenen Tags können allerdings von allen Nutzern verwendet werden, z.B. für die Suche nach einem Video.

Die zweite Art des Tagging umfasst das kollaborative Tagging oder Social Tagging [180], welches im Deutschen auch als *gemeinschaftliches Indexieren* bezeichnet wird [186]. Der Unterschied zum einfachen Tagging ist, dass die Zuweisung von Tags in der Regel durch mehrere Anwender stattfindet. Die Kategorisierung des Inhalts liegt also in den Händen aller Anwender und nicht nur im Einflussbereich einer einzigen Person. Die durch das Social Tagging erzeugte Klassifikation wird in der Literatur auch als *Folksonomy* bezeichnet und steht für eine Benutzer-generierte Taxonomie, die durch eine Einigkeit der Benutzer entsteht [57, S. 59]. Ein Beispiel für die im Social Tagging beteiligten Entitäten sowie deren Interaktion ist in Abbildung 3.6 dargestellt. Die Abbildung zeigt zwei Anwender, die einer Ressource frei definierte Tags zuweisen.

Beim Social Tagging wird die Wiederverwendung von Tags angestrebt, d.h., ein Anwender kann einen für eine Ressource vergebenen Tag ebenfalls der Ressource zuweisen. So lassen sich die Tags ermitteln, die am häufigsten für eine Ressource vergeben wurden und diese somit am besten beschreiben. Für die Darstellung von Tags wird oft eine sogenannte *Tag-Cloud* verwendet. Diese stellt die verwendeten Tags oft alphabetisch sortiert dar und hebt die am häufigsten genutzten Tags optisch hervor [186], z.B. durch einen größeren Schriftgrad.

In den vorigen Absätzen wurden als Hauptziele des Tagging die Kategorisierung und das Auffinden von Inhalten genannt. Im Rahmen des Social Tagging spielen jedoch auch weitere Gründe für das Tagging von Ressourcen eine Rolle. In der bestehenden Literatur wurden einige Motive des Tagging aus Sicht der Anwender genannt. Marlow et al. nehmen eine Unterteilung in organisatorische und soziale Motive vor [129]. Heckner et al. kategorisieren die Motive des Tagging ähnlich. Für sie handelt es sich bei den Motiven um eine persönliche Verwaltung von Informationen zur Wiederauffindbarkeit sowie um das Einstellen von Inhalten als Beitrag zur Gemeinschaft [65]. Eine etwas andere Klassifikation der Motivation beschreiben Körner et al. Für sie gibt es zum einen Anwender, die Inhalte kategorisieren und zum anderen Anwender, die Inhalte beschreiben [102]. In einer Studie zum Social Bookmarking Dienst *Delicious* identifizieren Golder und Huberman sieben verschiedene Rollen von Tags, welche Rückschlüsse auf die Motivation der Anwender ziehen lassen [53]. Basierend auf diesen sieben Arten von Tags leiten Sen et al. die folgenden drei Oberkategorien ab: Sachliche Tags, subjektive Tags und persönliche Tags [175]. Die Studien von Nov et al. sowie von Ames und Naaman betrachteten die Verwendung von Tags auf der Plattform Flickr¹. Beide Studien kategorisierten die Motive für das Tagging in zweidimensionaler Weise. Eine Dimension grenzt die Aspekte Organisation und Kommunikation ab, während die andere Dimension die Zielgruppe der Tags in den Anwender selbst sowie in die Gemeinschaft unterteilt [6, 141].

Wie aus der beschriebenen Literatur zu entnehmen ist, existiert keine einheitliche Motivation für die Erstellung von Tags. Dies deckt sich auch mit den Ergebnissen von Strohmaier et al., welche in ihrer Studie herausgefunden haben, dass sich die Motive für das Tagging zwischen Social Tagging Systemen, aber sogar auch innerhalb von solchen Systemen unterscheiden [188, S. 10]. Durch die unterschiedlichen Motive der Anwender unterscheidet sich auch die Art der erstellten Tags. Für den Rahmen dieser Arbeit wird eine Unterteilung der Tag-Arten in die von Sen et al. vorgeschlagenen Kategorien vorgenommen [175, S. 184 f.]:

- Persönliche Tags
- Sachliche Tags
- Subjektive Tags

Persönliche Tags werden von einem Anwender vergeben, um Inhalte für sich selbst zu klassifizieren, z.B. zur Organisation von Aufgaben anhand eines „toread“-Tags. Diese

¹<http://www.flickr.com> (besucht am 08.04.2014)

Art von Tag ist für andere Anwender in der Regel unbrauchbar und wird im Rahmen dieser Arbeit daher nicht näher betrachtet. Im Gegensatz dazu unterstützen sachliche und subjektive Tags eine soziale Interaktion. Mit Hilfe von sachlichen Tags werden einer Ressource Fakten zugewiesen. So kann beispielsweise der Name des Autors oder das Erscheinungsjahr des Buches als Tag verwendet werden. Dadurch lassen sich zum einen Ressourcen kategorisieren und zum anderen lassen sich auch verwandte Ressourcen leichter auffinden, z.B. Bücher desselben Autors. Diesbezüglich führen Xu et al. auch die sogenannten *Attribute Tags* auf, welche inhärente Attribute einer Ressource beschreiben, die nicht direkt aus dem Inhalt abgeleitet werden können [216].

Bei der dritten Tag-Kategorie nach Sen et al. handelt es sich um subjektive Tags. Tags dieser Art werden von Anwendern vergeben, um Gefühle und Meinungen bezüglich der jeweiligen Ressource zum Ausdruck zu bringen [53, 129, 175, 216], z.B. die Vergabe der Tags „lustig“ oder „sehenswert“ zu einem Film. Diese Formen von Tags zeigen also eine Art von Bewertungscharakteristik auf, die auch für Empfehlungsgenerierungen genutzt werden könnte. Wie die Studie von Sen et al. zeigt, können subjektive Tags aufgrund ihres Bewertungscharakters auch zur Entscheidungsfindung beitragen [175]. Die explizite Kombination von Tags und Online-Bewertungen stellt einen wichtigen Teil dieser Arbeit dar. Die wichtigsten verwandten Arbeiten hierzu werden in Abschnitt 4.3.2 noch weiter abgegrenzt.

Einen besonderen Verwendungszweck des Tagging stellen sogenannte *Hashtags* dar. Als Hashtag wird ein Schlagwort bezeichnet, dem ein Raute-Zeichen (engl. *hash symbol*) vorangestellt wird [79, S. 173]. Ein Beispiel hierzu wäre der Hashtag „#Berlin“. Hashtags werden typischerweise in Microblogging-Nachrichten verwendet, um die Teilnahme an bestimmten Gesprächsthemen zu signalisieren. Die bekannteste Plattform zur Verwendung von Hashtags ist der Microblogging-Dienst Twitter¹. Die auf dieser Plattform verbreiteten Nachrichten (sogenannte *Tweets*) können mit beliebigen Hashtags ausgezeichnet werden, wobei gewisse syntaktische Regeln für die Definition von Hashtags gelten, z.B. dürfen Hashtags keinerlei Leerzeichen enthalten. Die Popularität von Hashtags hat sich in der Vergangenheit wesentlich gesteigert. Dies zeigt sich unter anderem daran, dass auch das größte soziale Netzwerk Facebook² mittlerweile den Einsatz von Hashtags in Statusnachrichten ermöglicht.

3.5.2 Chancen und Risiken

Die Verlagerung der Kategorisierung von einer zentralen Instanz an eine Vielzahl von Anwender birgt sowohl Chancen als auch Risiken. Ein Vorteil ist, dass der Anbieter einer Webseite die Inhaltsklassifikation in die Hände der Nutzer legen kann, wodurch eine Kosteneinsparung möglich ist, z.B. weil kein Produktexperte benötigt wird, um die Klassifikation vorzunehmen. Zudem spiegeln die erzeugten Tags eher das Nutzungsverhalten wieder, da diese schließlich von Endnutzern und nicht von einzelnen Produktexperten ge-

¹<http://www.twitter.com> (besucht am 10.06.2014)

²<http://www.facebook.com> (besucht am 10.06.2014)

neriert wurden. Ferner können über die Zuweisung von Tags möglicherweise Rückschlüsse auf die Vorlieben von Anwendern gezogen werden. Vergibt beispielsweise ein Kunde für ein Produkt einen bestimmten Tag, welcher auch von einem anderen Anwender für ein anderes Produkt verwendet wurde, so kann hier eine Verbindung hergestellt werden, welche möglicherweise als Basis für Produktempfehlungen verwendet werden kann.

Wie bereits zuvor erwähnt wurde, liegt die Wahl der verwendeten Tags komplett bei den Anwendern. Es stellt sich daher die Frage, ob Anwender ohne eine zentrale Instanz überhaupt in der Lage sind eine sinnvolle Kategorisierung des Inhalts vorzunehmen. Zu diesem Zweck wurde von Halpin et al. eine Studie zu den 500 populärsten Lesezeichen des Social Bookmarking Dienstes Delicious¹ durchgeführt. Die Ergebnisse der Studie zeigen eine Verteilung die Potenzgesetzen (engl. *power laws*) folgt, d.h., die betrachteten Tags haben sich für die untersuchten Lesezeichen über die Zeit stabilisiert und die Anwender haben letztlich einen Konsens über die am besten geeigneten Tags erreicht [59]. Dies zeigt, dass Anwender auch ohne zentrale Instanz in der Lage sind eine sinnvolle Inhaltskategorisierung vorzunehmen.

Die fehlende zentrale Instanz in Tagging Systemen wirkt sich allerdings trotzdem negativ auf die Qualität aus, denn es gibt meist keine Möglichkeit das verwendete Vokabular der Anwender zu kontrollieren oder zu korrigieren. Dies kann auf verschiedene Arten zu einer Qualitätsminderung des Tagging Systems führen. Ein großes Problem ist die Mehrdeutigkeit von Begriffen (Tags), z.B. in Form von Synonymen oder Homonymen, aber auch die Verwendung der gleichen Begriffe im Singular und im Plural ist problematisch [57]. Ferner reduzieren Rechtschreibfehler in Tags die Qualität des Tagging Systems. Aber auch die in Abschnitt 3.5.1 beschriebenen persönlichen Tags (wie z.B. „toread“) verringern den Nutzen des Tagging Systems, da diese keinen Mehrwert für andere Nutzer darstellen, aber die Anzahl der Tags erhöhen und dadurch die Übersichtlichkeit verringern. Ein weiteres Problem stellt das Spamming von Tags dar, bei dem Anwender (oder automatisierte Mechanismen) Tags erzeugen, die dem Ersteller Vorteile verschaffen sollen, z.B. in Form von Werbung. Letztlich leiden Tagging Systeme auch am Kaltstartproblem, bei dem eine neu angelegte Ressource anhand von Tags nicht bzw. schwer auffindbar ist, da noch keine Tags dazu existieren.

Zu den oben genannten Problemen existieren in der Literatur verschiedene Lösungsansätze. Pind nennt einige Ansätze zur Verbesserung der Tag-Qualität, z.B. das Vorschlagen von existierenden Tags anderer Anwender oder die automatische Erkennung von Synonymen [156]. Bezüglich des Spam-Problems schlagen Xu et al. einen Ansatz vor, der die Bewertung von Tags sowie die Reputation der einzelnen Anwender umfasst [216]. Ferner beschreiben Xu et al., ähnlich zu Pind, eine Inhalts-basierte Generierung von Tag-Vorschlägen zur Lösung des Kaltstartproblems [216]. Ein weiterer Ansatz zur Verbesserung der Tag-Qualität ist auf der Webseite von Librarything² zu beobachten. Dort wird Anwendern ein Votingsystem bereitgestellt, in dem diese für die Zusammenlegung von semantisch ähnlichen Tags abstimmen können.

¹<http://www.delicious.com> (besucht am 07.04.2014)

²http://www.librarything.com/wiki/index.php/Tag_combining (besucht am 08.04.2014)



Abbildung 3.7: Einordnung des Themas im Bereich Social Media

3.6 Zusammenfassung

In diesem Kapitel wurden die für diese Arbeit relevanten verwandten Arbeiten aus dem Bereich Social Media näher betrachtet. Eine große Bedeutung kommt hier dem Social Commerce zu, welches den Einsatz von Social Media im E-Commerce beschreibt. Ferner wurden die momentan in B2C-Onlineshops genutzten Social Media Features analysiert. Die drei Themenbereiche Social Tagging, Produktbewertungen und Produktrezensionen wurden dann detailliert betrachtet und abgegrenzt. Abschließend wurde die bestehende Forschungsliteratur zur Verknüpfung von Online-Bewertungen und Social Tagging aufgeführt. Die Kombination von Social Tagging, Produktbewertungen und Produktrezensionen ist auch gleichzeitig der Schwerpunkt des nächsten Kapitels. Die Einordnung der Arbeit in diese Themenbereiche sowie in den übergeordneten Themenkomplex des Social Media ist in Abbildung 3.7 nochmal grafisch dargestellt.

Kapitel 4

Konzeptionelle Betrachtung von Rated Tags

In diesem Kapitel wird zunächst das Forschungsproblem dieser Arbeit konkretisiert. Danach erfolgt die Präsentation eines eigenen Lösungsansatzes mit der Bezeichnung *Rated Tags*. Dabei wird der fachliche Rahmen des Ansatzes erläutert, welcher unter anderem die beteiligten Akteure und Entitäten sowie relevante Anwendungsfälle umfasst. Abschließend erfolgt eine Abgrenzung des Forschungsansatzes zu bestehenden praktischen Lösungsansätzen sowie zu verwandten Forschungsarbeiten.

4.1 Probleme aktueller Bewertungssysteme

Die in Abschnitt 3.4 erläuterten Konzepte der Produktbewertung und Produktrezension werden nachfolgend unter dem Begriff *Bewertungssystem* zusammengefasst. Wie sich in Abschnitt 3.4 gezeigt hat, haben aktuelle Bewertungssysteme jeweils Stärken und Schwächen. Die Stärke von Produktbewertungen ist die Einfachheit der Erstellung sowie eine automatisierte Verarbeitung, z.B. zur Berechnung der Durchschnittsbewertung. Dies sorgt für einen geringeren kognitiven Aufwand bei den Anwendern. Auf der anderen Seite sind die Gesamtbewertungen für Produkte oder Dienstleistungen nur ein grober Qualitätsindikator. Die speziellen Anforderungen von Konsumenten lassen sich über die Gesamtbewertung alleine jedoch nicht überprüfen. Hierzu eignen sich Produktrezensionen. Diese enthalten oft sehr detaillierte Informationen und Meinungen zu den einzelnen Merkmalen eines Produkts. Ein Problem aktueller Bewertungssysteme ist allerdings, dass die Anzeige dieser Rezensionen oft nur sortiert nach Datum stattfindet und eine Filterung anhand der speziellen Konsumentenanforderungen nicht möglich ist. Da die Analyse aller Rezensionen zeitlich oft nicht möglich bzw. sinnvoll ist, begnügt sich ein Konsument häufig mit einer Untermenge an Rezensionen. Dadurch können womöglich nicht alle relevanten Informationen erfasst werden, was zu einer geringeren Entscheidungsqualität führen kann [121, S. 235]. Zudem sind Produktrezensionen aufgrund ihrer textuellen Form schwer automatisiert zu verarbeiten, wodurch für den Anwender ein hoher kognitiver Aufwand entsteht, da viele Produktrezensionen gelesen und die darin enthaltenen Informationen zu den potenziellen Produkten miteinander verglichen werden müssen. Die beschriebene Problematik wird anhand des nachfolgend dargestellten Beispielszenarios verdeutlicht.

Beispielszenario Der Konsument Bob sucht einen Full HD Beamer für die Einrichtung eines Heimkinos. Gemäß der ersten Phase des Entscheidungsfindungsprozesses (siehe Seite 18) definiert Bob zunächst seine Anforderungen an das Produkt. Da Bob sehr

8 von 20 Kunden fanden die folgende Rezension hilfreich

☆☆☆☆ **Serie Super, aber...**, 19. Oktober 2013

Von [CSXdesign](#) (Freiberg) - [Alle meine Rezensionen ansehen](#)

Rezension bezieht sich auf: *Game of Thrones - Die komplette erste Staffel [5 DVDs] (DVD)*

...unnötiger Kopierschutz verhindert PC-Wiedergabe.

Er sorgt dafür, dass die DVDs nicht mit einem PC-Laufwerk abgespielt werden kann.

WindowsMediaPlayer ruckelt dann stark und über den Bildern liegen grüne Texturen, sodass ein gucken der DVD unmöglich wird. Wenn der Player dann abstürzt kommt auch die Meldung, dass es am Kopierschutz der DVD liegt.

Abbildung 4.1: Beispiel für die geringe Aussagekraft einer Gesamtbewertung (Quelle: <http://www.amazon.de>, besucht am 10.02.2014)

geräuschempfindlich ist, wählt er als wichtigste Anforderung ein geringes Betriebsgeräusch des Beamers. Im nächsten Schritt besucht Bob einen Onlineshop, um nach potenziellen Beamern zu suchen und diese zu analysieren. Als grober Indikator dient Bob zunächst die Gesamtbewertung der Produkte. Aus dieser Bewertung geht allerdings nicht hervor, wie der jeweilige Beamer im Hinblick auf das Kriterium Betriebsgeräusch abschneidet. Bob könnte alternativ auch die Produktblätter der Hersteller zu den einzelnen Beamern studieren, um die dort genannten Werte zu den Betriebsgeräuschen zu vergleichen. Allerdings vertraut Bob auf die subjektiven Eindrücke anderer Konsumenten und nicht auf die von den Herstellern genannten Eigenschaften. Daher ist Bob auf die Analyse der Kundenrezensionen der einzelnen Produkte angewiesen. Für häufig gekaufte Produkte existieren teilweise mehrere hundert Rezensionen, wodurch eine Analyse aller Rezensionen zu aufwändig wäre. Hinzu kommt, dass Bob vor dem Lesen einer Rezension nicht weiß, ob diese überhaupt den Aspekt des Betriebsgeräusches adressiert.

Das obige Beispielszenario hat verdeutlicht, dass eine Lücke zwischen Produktbewertungen und Produktrezensionen besteht, welche für den zeitlichen und kognitiven Aufwand von Konsumenten verantwortlich ist. Ein weiteres Problem, welches im obigen Beispiel nicht betrachtet wurde, ist neben der geringen Aussagekraft von Gesamtbewertungen, auch deren mögliche Fehlinterpretation. Für die Verdeutlichung dieses Sachverhalts wird auf Abbildung 4.1 verwiesen. Die Abbildung zeigt eine Kundenrezension für eine Serie. Wie der vergebene Kurztext „*Serie Super, aber...*“ vermuten lässt, ist der Rezensent offensichtlich von der Serie begeistert, vergibt jedoch nur einen von fünf möglichen Sternen. Der Grund hierfür ist nur in der Rezension zu finden. Aus einer alleinigen Betrachtung der Gesamtbewertung würden andere Konsumenten möglicherweise falsche Rückschlüsse ziehen.

4.2 Eigener Lösungsansatz

Wie der vorige Abschnitt gezeigt hat, existiert in aktuellen Bewertungssystemen eine Lücke zwischen Gesamtbewertungen und Rezensionen, welche den Entscheidungsfindungsprozess

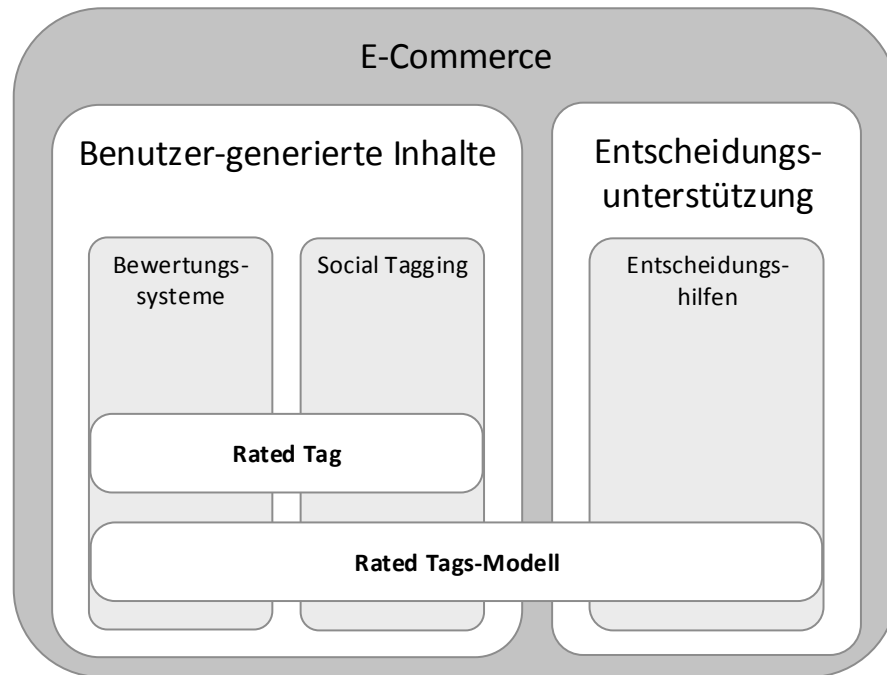


Abbildung 4.2: Einordnung des Rated Tags-Modells in verwandte Themenbereiche

von Konsumenten erschwert. Weiterhin hat die Anwenderstudie von Ozakca und Lim zur Nutzung von Rezensionen und Gesamtbewertungen gezeigt, dass Anwender gerne die verschiedenen positiven und negativen Aspekte innerhalb Rezension vergleichen und dass hierzu eine Bewertung von einzelnen Aspekten sinnvoll wäre [146, S. 1186].

Die grundlegende Idee des eigenen Lösungsansatzes ist es, Anwendern eine Möglichkeit bereitzustellen, um die in einer Rezension beschriebenen Aspekte separat zu spezifizieren und zu bewerten. Hierfür wurde das Prinzip des Tagging in Betracht gezogen, welches in Abschnitt 3.5 bereits näher erläutert wurde. Tagging bietet die Möglichkeit einer Ressource Schlagwörter zuzuweisen, um diese leichter auffindbar zu machen. Für den dargestellten Lösungsansatz wird Tagging dazu verwendet, um den Anwendern die Möglichkeit zu geben, die für eine Bewertung relevanten Produktmerkmale einer Rezension zuzuweisen. Im Sinne des Tagging stellt eine Rezension hier also die jeweilige Ressource für die Verschlagwortung dar.

Zusätzlich zum normalen Tagging ist jedoch für den Ansatz auch die Bewertung dieser Tags, bei denen es sich um die jeweiligen Produktmerkmale handelt, notwendig. Aus dieser Anforderung wurde ein hybrider Tag konzipiert, bei dem es sich um eine Kombination aus Tag und Bewertungsskala handelt. Diese spezielle Form eines Tags wird nachfolgend als *Rated Tag* bezeichnet [92]. Ein *Rated Tag* ist also ein Benutzer-generiertes Schlagwort, welches zusätzlich eine Bewertungskomponente umfasst. Der nachfolgend präsentierte Lösungsansatz basiert auf *Rated Tags* und wird daher als *Rated Tags-Ansatz* oder *Rated Tags-Modell* bezeichnet. Eine Einordnung des *Rated Tags-Modells* in verwandte Themenbereiche ist Abbildung 4.2 zu entnehmen.

Die nachfolgenden Ausführungen des fachlichen Rahmens legen eine möglichst generische Betrachtungsweise zu Grunde. Daher wird durchgehend der Begriff „Ressource“ verwendet, um Objekte zu adressieren, welche mit Rated Tags gekennzeichnet werden können. Hierzu zählen beispielsweise Produkte, Dienstleistungen oder andere Objekte, die bewertbare Merkmale enthalten. Die abstrakte Bezeichnung „Ressource“ wird in dieser Form auch als Entität im Rahmen eines traditionellen Tagging Systems verwendet (siehe Abschnitt 3.5.1).

4.2.1 Ziel

Das übergeordnete Ziel des Rated Tags-Ansatzes ist die Verbesserung des Entscheidungsfindungsprozesses von Konsumenten. Dieses Ziel kann weiter unterteilt werden in die Ziele „Reduzierung des Entscheidungsaufwandes“ und die „Erhöhung der Entscheidungsqualität“.

Der Entscheidungsaufwand eines Konsumenten basiert primär auf dem zeitlichen Aufwand, d.h. auf der Zeit die ein Konsument benötigt, um Produkte oder Dienstleistungen hinsichtlich der eigenen Anforderungen zu analysieren, miteinander zu vergleichen und letztendlich eine geeignete Auswahl zu treffen. Abhängig von der Präsentation der Informationen spielt ebenfalls der kognitive Aufwand der Konsumenten eine Rolle. Weisen die präsentierten Informationen eine geringe Strukturierung auf oder sind diese für die Konsumenten schwer zu verarbeiten, so steigt der kognitive Aufwand des Konsumenten. Das Ziel des Rated Tags-Systems ist die Reduzierung der genannten Aufwände für Konsumenten.

Ein weiteres Ziel des Rated Tags-Systems ist die Erhöhung der Entscheidungsqualität der Konsumenten. Die Qualität einer Entscheidung ist unter anderem dadurch definiert, inwieweit die von den Konsumenten geforderten Anforderungen mit den Charakteristiken des Produkts bzw. der Dienstleistung übereinstimmen. Traditionelle Bewertungssysteme, die nur Gesamtbewertungen und Rezensionen umfassen, bieten den Konsumenten diesbezüglich nur eine begrenzte Entscheidungsunterstützung an. Ziel des Rated Tags-Systems ist es daher, den Konsumenten eine Möglichkeit an die Hand zu geben, um entscheidungsrelevante Ressourcen hinsichtlich ihrer Anforderungen leichter zu finden und auszuwerten.

4.2.2 Beispielszenario

Um das Konzept und den Mehrwert des Rated Tags-Ansatzes besser zu verdeutlichen, wird in Abbildung 4.3 ein beispielhaftes Szenario zur Verwendung von Rated Tags aufgezeigt. Die Abbildung zeigt die vier Komponenten Anwender, Rated Tags, Rezensionen und Produkt. Das Szenario aus Abbildung 4.3 zeigt zwei Anwender, die jeweils eine Rezension für ein Produkt erstellen. Beide Anwender fügen ihrer Rezension Rated Tags hinzu, um gezielt Produktmerkmale samt Bewertung zu spezifizieren. Die Rezensionen, und damit auch die dafür vergebenen Rated Tags, sind jeweils genau einem Produkt zugewiesen.

Der eigentliche Mehrwert von Rated Tags ergibt sich dadurch, dass andere Anwender die vergebenen Rated Tags für ihre Entscheidungsfindung verwenden können. Dies

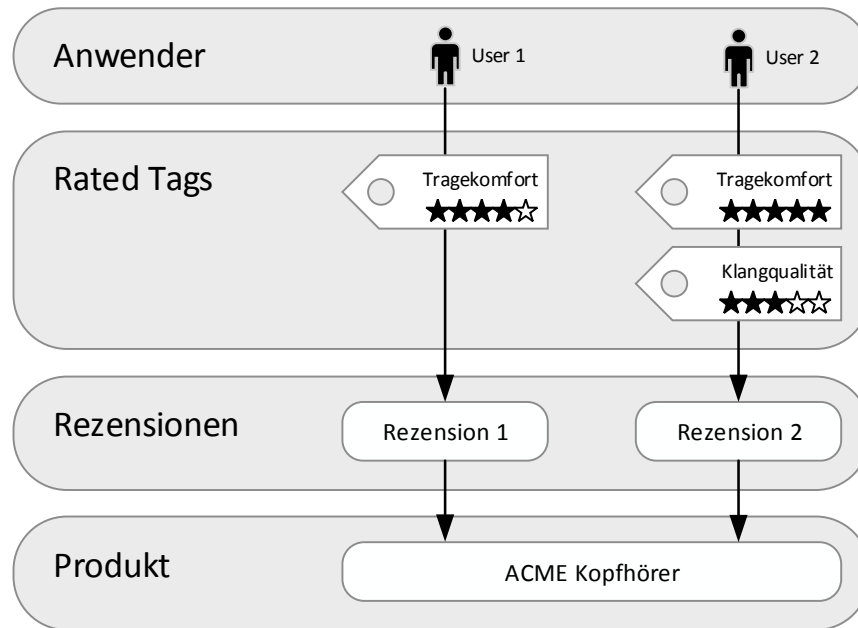


Abbildung 4.3: Beispielhafte Darstellung des Rated Tags-Ansatzes

wird beispielhaft in Abbildung 4.4 dargestellt. Die Abbildung vergleicht anhand eines Beispielszenarios („Ein Konsument sucht nach einem Kopfhörer mit hohem Tragekomfort“) die Entscheidungsfindung eines traditionellen Bewertungssystems und des Rated Tags-Systems. Anwender eines traditionellen Bewertungssystems sehen nur eine Gesamtbewertung und müssen alle Rezensionen lesen, um die Meinungen zum Tragekomfort der Kopfhörer herauszufinden. Im Fall des Rated Tags-Systems wird das Produktmerkmal „Tragekomfort“ durch einen Rated Tag repräsentiert. Anwender sehen so zum einen die Durchschnittsbewertung für das Merkmal Tragekomfort und zum anderen haben sie die Möglichkeit nur Rezensionen anzuzeigen, welche mit dem Rated Tag „Tragekomfort“ gekennzeichnet wurden.

4.2.3 Akteure

Als Akteure lassen sich grundsätzlich die Anwender und Anbieter des Rated Tags-Systems identifizieren. Als potenzielle Anbieter des vorgestellten Rated Tags-Konzepts kommen Webseiten infrage, welche die Benutzer-generierte Erzeugung von Inhalten in Form von Kundenmeinungen und Bewertungen für spezielle Ressourcen zulassen. Eine Ausprägung hiervon sind Onlineshops, wobei hier der Schwerpunkt eher im B2C-E-Commerce liegt, da in diesem Modell Endkonsumenten angesprochen sind, welche im Vergleich zu Businesskunden ihre Erfahrungen häufiger über Kundenmeinungen (Rezensionen) austauschen.

Neben Onlineshops ist auch die Nutzung innerhalb von Bewertungsportalen denkbar. Als Bewertungsportal werden alle Webseiten klassifiziert, die Anwendern die Bewertung

Entscheidungsfindung

Szenario: Ein Kunde sucht einen komfortablen Kopfhörer

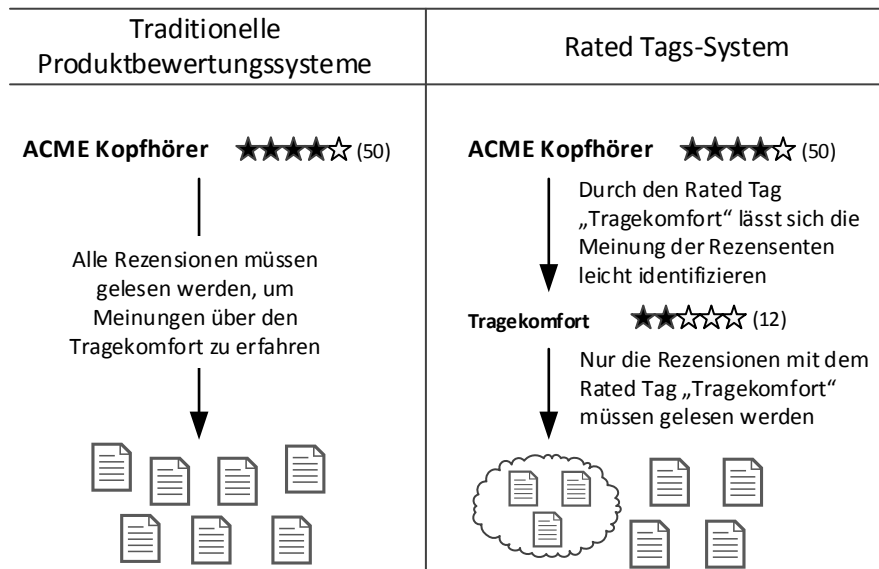


Abbildung 4.4: Beispielszenario zur Verwendung von Rated Tags, um die Entscheidungsfindung zu verbessern

von bestimmten Ressourcen (z.B. Produkte, Dienstleistungen o.ä.) erlauben. Eine direkte Verkaufsabsicht besteht im Gegensatz zu Onlineshops hier meist nicht. Bekannte Bewertungsportale für Produkte sind beispielsweise Ciao¹ oder Epinions². Aber auch Bewertungsportale aus dem Bereich Tourismus könnten von Rated Tags profitieren. Ein bekanntes Bewertungsportal ist TripAdvisor³. Im Vergleich zu Produktmerkmalen könnten im Bereich Tourismus beispielsweise spezielle Hotelmerkmale mittels Rated Tags erfasst werden. Wie die möglichen Einsatzbereiche zeigen, ist das entworfene Rated Tags-Modell so konzipiert, dass es unabhängig von der jeweiligen Anbieterdomäne generisch eingesetzt werden kann.

In Anlehnung an die oben genannten Anbieter können die Anwender des Rated Tags-Ansatzes in die Gruppen „Inhaltsproduzent“ und „Inhaltskonsument“ untergliedert werden. Inhaltsproduzenten nutzen die bereitgestellte Funktionalität zur Erstellung von Rezensionen und Rated Tags, um ihr Wissen bzw. ihre Erfahrung in Bezug auf eine Ressource (z.B. ein Produkt) mit anderen zu teilen. Die Anwendergruppe der Inhaltskonsumenten hingegen sind auf der Suche nach Informationen, um ein Entscheidungsproblem zu lösen (z.B. „Welches Hotel eignet sich für meine Bedürfnisse am besten?“). Inhaltskon-

¹<http://www.ciao.de> (besucht am 18.05.2014)

²<http://www.epinions.com> (besucht am 18.05.2014)

³<http://www.tripadvisor.com> (besucht am 18.05.2014)

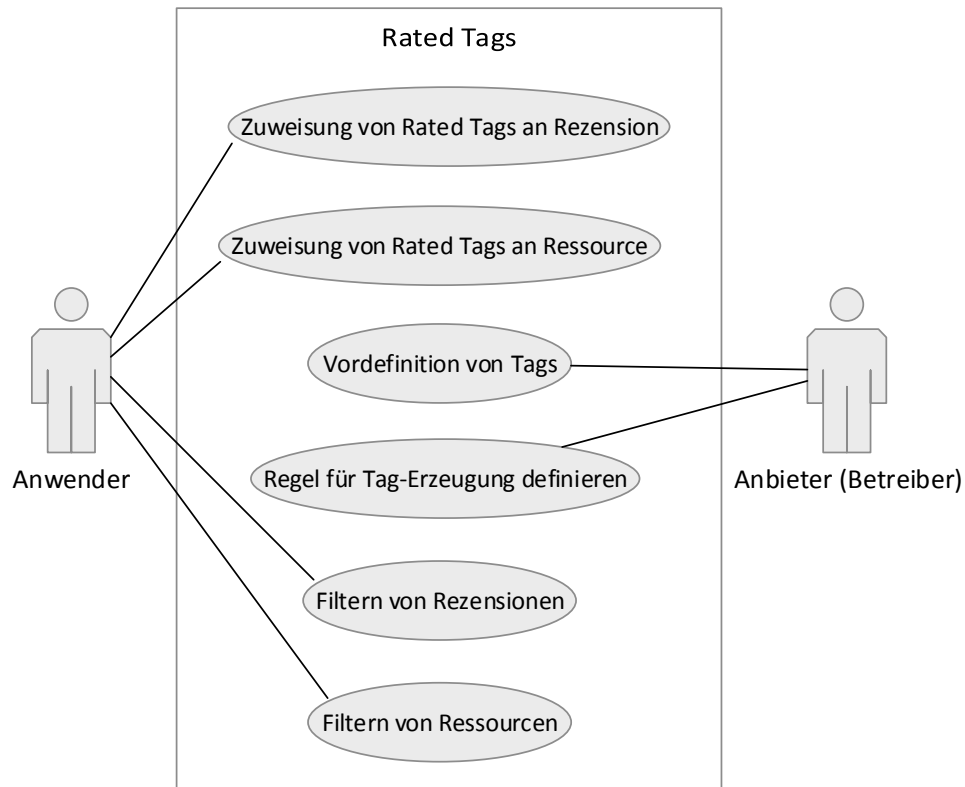


Abbildung 4.5: UML-Anwendungsfalldiagramm für das Rated Tags-Modell

sumenten nutzen Rezensionen und Rated Tags, um passende Ressourcen zu ermitteln und hinsichtlich der eigenen Anforderungen zu vergleichen.

4.2.4 Anwendungsfälle

Die Anwendungsfälle des Rated Tags-Systems lassen sich grob in drei Kategorien unterteilen. Zum einen existieren Anwendungsfälle die eine Erzeugung und Zuweisung von Rated Tags an Ressourcen betreffen. Ferner sind Anwendungsfälle identifizierbar, welche sich der Auswertung der Rated Tags widmen, um den Entscheidungsfindungsprozess von Konsumenten zu verbessern. Die letzte Kategorie betrifft die Verwaltung von Rated Tags durch den Anbieter. Ein Überblick der einzelnen Anwendungsfälle ist in Abbildung 4.5 zu finden. Die in der Abbildung dargestellten Anwendungsfälle werden gemäß *Unified Modeling Language (UML)* als Anwendungsfalldiagramm dargestellt und nachfolgend näher beschrieben.

Erzeugung von Rated Tags

Bei der Erzeugung von Rated Tags lassen sich prinzipielle zwei Anwendungsfälle unterscheiden, nämlich die Erzeugung und Zuweisung von Rated Tags an Rezensionen

sowie die Erzeugung und Zuweisung an Ressourcen. Bei Ersterem schreibt ein Anwender eine entsprechende Rezension und vergibt hierfür passende Rated Tags. Der zweite Fall der Tag-Erzeugung betrifft die Zuweisung von Rated Tags an Ressourcen. Hier sollen Anwendern auch ohne die Erstellung einer textuellen Rezension eine Bewertung in Form von Rated Tags vornehmen können. Eine Zuweisung von einfachen Tags an eine Ressource ist im Rated Tags-Modell nicht vorgesehen, da sich dieser Anwendungsfall nicht von einem traditionellen Tagging Systems unterscheidet. Es ist anzumerken, dass das einfache Tagging in erster Linie der Kategorisierung dient, während sich die Ansätze des Rated Tags-Modells eher mit den Merkmalen von Ressourcen (z.B. Bildqualität etc.) beschäftigen.

Durch die Erzeugung von Rated Tags für Ressourcen und Rezensionen entsteht eine Menge an Ressourcenmerkmalen (inkl. durchschnittlicher Bewertung). Die Übersicht aller genannten Ressourcenmerkmale kann für Konsumenten hilfreich sein, um zu entscheiden, ob und welche Merkmale für die Entscheidungsfindung eine Rolle spielen. Es ist nämlich nicht immer so, dass Konsumenten bereits vor dem Lesen der Rezensionen genau abgegrenzte Kriterien formuliert haben. Vielmehr entdecken Konsumenten erst durch das Lesen von Rezensionen wichtige Ressourcenmerkmale, die sie bislang nicht in die Entscheidungsfindung miteinbezogen haben [121, S. 232]. Durch die Darstellung aller genannten Ressourcenmerkmale wird diese Identifikation der relevanten Merkmale vereinfacht.

Das Modell sieht vor, dass die Erstellung von Rated Tags nur Anwendern ermöglicht wird. Dem Anbieter des Systems wird die Erstellung vorenthalten, um Manipulationen zu erschweren. Als weiterer Mechanismus gegen Manipulation und Spam wurde die Einführung eines Rollensystems in Betracht gezogen, so dass nur Anwendern ab einer gewissen Reputationsstufe Rated Tags erstellen dürfen. Diese Idee wurde jedoch vorerst nicht in das Modell mitaufgenommen, weil davon auszugehen ist, dass eine solche Maßnahme die Benutzbeteiligung negativ beeinflusst und somit den Nutzen des Systems reduzieren würde.

Analyse von Tags

Neben der Erstellung von Rated Tags liegt das Hauptaugenmerk der Funktionalität in der Auswertung der Rated Tags, um den Entscheidungsfindungsprozess von Anwendern zu unterstützen. Im Gegensatz zu passiven Recommender Systemen erlaubt der Rated Tags-Ansatz eine explizite Benutzerinteraktion und fällt damit in die Kategorie der interaktiven Entscheidungshilfen. Die wesentlichen entscheidungsunterstützenden Funktionen des Rated Tags-Systems werden nachfolgend erläutert.

Eine entscheidungsunterstützende Funktion des Rated Tags-Ansatzes umfasst das Filtern von Ressourcen basierend auf den vorhandenen Tags. Dies wird in Abbildung 4.5 durch den Anwendungsfall „Filtern von Ressourcen“ ausgedrückt. Da Rated Tags direkt oder indirekt einer Ressource zugeordnet werden, können Filtermechanismen genutzt werden, um anhand von Rated Tags relevante Ressourcen leichter zu ermitteln. Sind beispielsweise mehrere Ressourcen mit dem Rated Tag „Bildqualität“ versehen, so könnte der Anwender die Suche auf die Ressourcen einschränken, deren Rated Tag „Bildqualität“

mindestens 4 von 5 Punkten erhalten hat. Die Definition von solchen Schwellenwerten (engl. *thresholds*) hilft Anwendern die konjunktive Entscheidungsstrategie anzuwenden, welche bereits in Abschnitt 2.2.3 näher erläutert wurde. Die beschriebene Funktionalität kann von Anwendern dazu eingesetzt werden, um das sogenannte *consideration set*, also die potenzielle Anzahl an relevanten Ressourcen, zu reduzieren. Ferner lässt sich durch den Zugriff auf die vergebenen Rated Tags auch eine Sortierung abbilden, z.B. um Ressourcen in Abhängigkeit eines Rated Tags in auf- oder absteigender Reihenfolge anzuzeigen.

Ein weiterer wichtiger Anwendungsfall ist das in Abbildung 4.5 dargestellte „Filtern von Rezensionen“. Dieser Anwendungsfall bezieht sich auf die Zuweisung von Rated Tags an Rezensionen. Die vergebenen Tags, welche letztendlich Merkmale der Ressource repräsentieren, werden dabei mit Rezensionen verknüpft. Dadurch können sich Anwender nur diejenigen Rezensionen anzeigen lassen, welche bestimmte Merkmale adressieren. Ist ein Konsument beispielsweise auf der Suche nach einem komfortablen Kopfhörer, so sucht er speziell nach Rezensionen die dieses Merkmal (hier „Tragekomfort“) beschreiben. Traditionelle Bewertungssysteme unterstützen Konsumenten diesbezüglich nur wenig. Aus der Vielzahl an Rezensionen werden den Konsumenten oft nur nach Datum sortierte oder als hilfreich eingestufte Rezensionen zuerst präsentiert. Da Konsumenten meist nur einen Teil aller Rezensionen analysieren können, besteht die Gefahr, dass Konsumenten basierend auf diesen Rezensionen eine suboptimale Entscheidung treffen, da möglicherweise nicht alle relevanten Informationen berücksichtigt wurden [121, S. 235]. Dies wurde auch durch das Effort-Accuracy Framework von Johnson und Payne festgehalten [87]. Das Filtern von Rezensionen soll Anwendern nur relevante Rezensionen anzuzeigen und somit die Komplexität und auch die benötigte Zeit für die Entscheidungsfindung verringern. Im Vergleich zum vorher genannten Anwendungsfall ist das Filtern von Rezensionen primär zur Analyse einer einzelnen Ressource und nicht zur Selektion von mehreren Alternativen ausgelegt.

Um zu demonstrieren an welchen Stellen im Entscheidungsfindungsprozess die oben genannten Anwendungsfälle die Konsumenten unterstützen können, wurde ein UML-Aktivitätsdiagramm erstellt, welches Abbildung 4.6 zu entnehmen ist. Das Diagramm zeigt einen beispielhaften Entscheidungsfindungsprozess von der Problemdefinition bis hin zur Entscheidung. Die schattierten Aktionen illustrieren, in welchen Phasen des Entscheidungsfindungsprozesses der Rated Tags-Ansatz die Konsumenten unterstützen soll.

Abbildung 4.6 gibt weiterhin Aufschluss über die Anwendergruppe die primär von Rated Tags profitieren könnte. Die schattierten Aktionen sind vornehmlich für Konsumenten interessant, welche detaillierte Anforderungen an eine Ressource stellen und Rated Tags zur Überprüfung dieser Anforderungen verwenden. Für Konsumenten, die beispielsweise eine schnelle Entscheidung treffen müssen oder deren Entscheidung nicht von einzelnen Merkmalen (z.B. nur vom Preis) abhängig ist, ist der Einsatz von Rated Tags möglicherweise nicht erforderlich. Für diese Fälle kann sich eine Entscheidung bereits nach der ersten oder zweiten Abzweigung (siehe Abbildung 4.6) ergeben. Ein Beispiel wäre die Sortierung der Alternativen nach dem Kriterium Preis und eine Selektion der günstigsten Alternative.

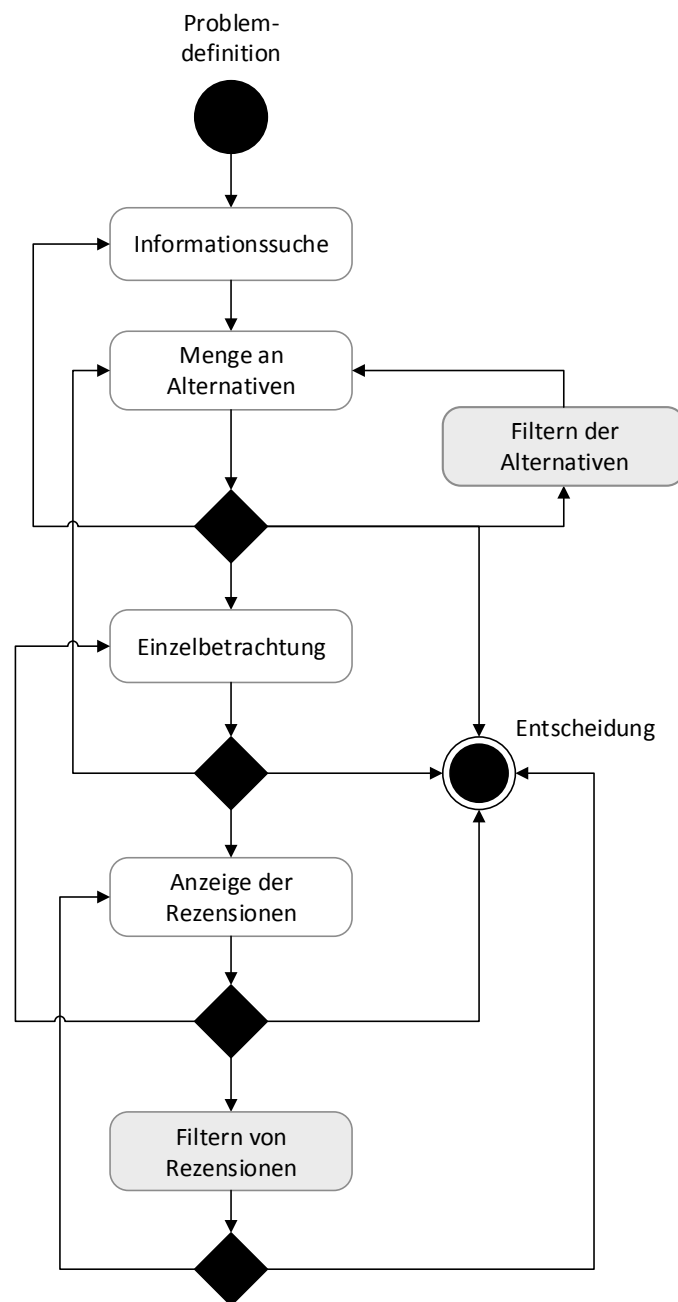


Abbildung 4.6: UML-Aktivitätsdiagramm zur beispielhaften Darstellung eines Entscheidungsprozesses. Die schattierten Aktionen illustrieren, in welchen Phasen des Entscheidungsfindungsprozesses der Rated Tags-Ansatz die Konsumenten unterstützen soll.

Verwaltung von Tags

Die Verwaltung von Tags betrifft die Anbieterseite des Systems. Wie erwähnt, sind Betreiber nicht in der Lage Rated Tags zu erstellen, um Manipulationen von Rated Tag-Bewertungen zu vermeiden. Jedoch wird es ihnen ermöglicht Ressourcen-typische Tags vorzudefinieren. Dies ist in Abbildung 4.5 als Anwendungsfall „Vordefinition von Tags“ dargestellt. Ein Bewertungsportal für Hotels könnte beispielsweise die Tags „Lage“, „Sauberkeit“ oder ähnliche Tags vordefinieren. Dies reduziert zum einen das Kaltstartproblem, das Tagging Systeme typischerweise mit sich bringen, indem die vordefinierten Tags von Anfang an zur Verfügung gestellt werden. Zum anderen hat der Betreiber direkten Einfluss auf die aus seiner Sicht wichtigsten Tags und kann somit die Qualität des Systems verbessern.

Ein weiterer Anwendungsfall für Betreiber ist die optionale Definition von Regeln für die Erzeugung von Tags. Damit soll Betreibern ermöglicht werden, die Qualität der erstellten Tags zu regulieren, indem Benutzer-generierte Tags vor ihrer Speicherung gegen diese Regeln geprüft und bei Nichtkonformität abgelehnt werden. Durch die Definition von bestimmten Regeln könnte beispielsweise die Erzeugung von unerwünschten Tags, z.B. Tags die Schimpfwörter enthalten, reduziert werden.

Weitere Anforderungen

Aufgrund der in den obigen Abschnitten beschriebenen Funktionalität zur Sortierung und Filterung von Ressourcen sowie der interaktiven Nutzung durch Anwender, kann das Rated Tags-System als interaktives Informationsverwaltungs-Tool (IIMT) klassifiziert werden [22, S. 226] [58, S. 163]. Basierend auf der Literatur aus den Bereichen der Informationssysteme und Entscheidungsunterstützung, leitet Pfeiffer die folgenden sechs Anforderungen an ein IIMT ab [155, S. 114]:

- **Adaptivität:** Die Benutzerschnittstelle muss sich adaptiv an die Bedürfnisse des Anwenders anpassen, um verschiedene Anwendergruppen (z.B. Anfänger und Experten) angemessen zu unterstützen.
- **Flexibilität:** Die Anwender müssen in der Lage sein die Interaktion jederzeit zu beenden. Es sollte eine Modularisierung vorhanden sein, damit Anwender flexibel verschiedene Entscheidungsstrategien anwenden können.
- **Geringer Aufwand:** Die Benutzerschnittstelle muss intuitiv bedienbar sein und den kognitiven Aufwand der Anwender minimieren.
- **Konsistenz:** Die Ergebnisse der Benutzerinteraktion müssen in konsistenter Weise bereitgestellt werden.
- **Steuerung:** Die Steuerung der Interaktion muss ganz beim Anwender liegen.
- **Transparenz:** Die Aktionen des IIMT müssen für die Anwender nachvollziehbar und transparent sein. Der Effekt einer Aktion muss den Anwendern möglichst unverzüglich angezeigt werden.

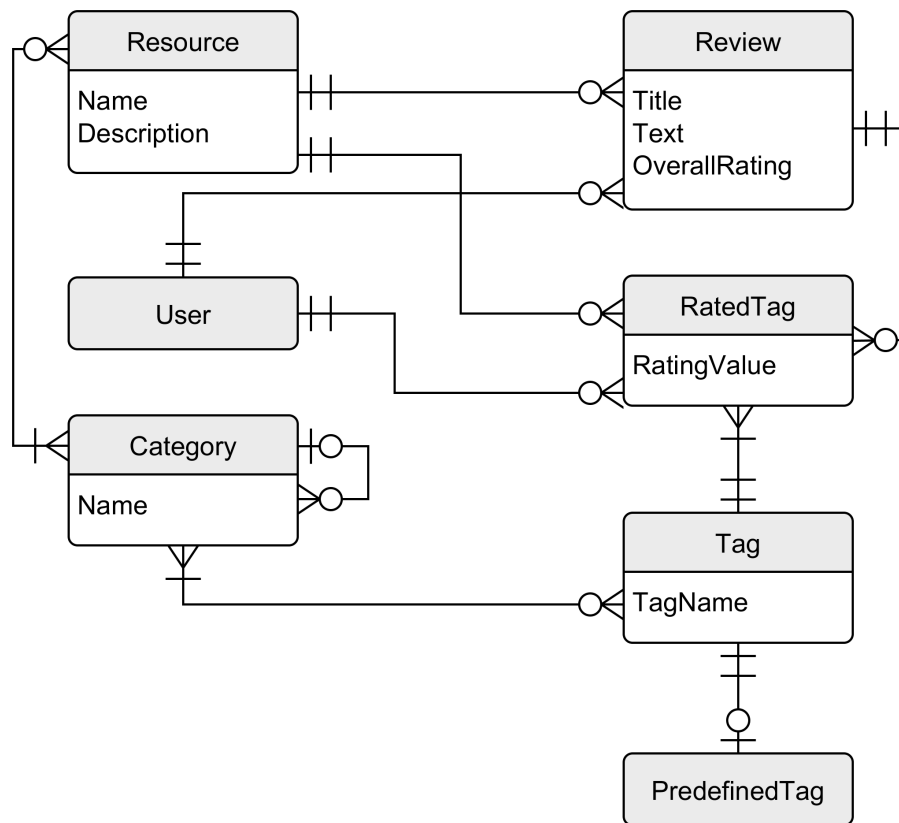


Abbildung 4.7: Vereinfachte Darstellung der relevanten Entitäten und Beziehungen des Rated Tags-Modells in Form eines ER-Diagramms

Die von Pfeiffer genannten nichtfunktionalen Anforderungen müssen für die oben genannten Anwendungsfälle berücksichtigt werden. Die konkrete Umsetzung dieser Anforderungen wird im Entwurf der Benutzeroberfläche (siehe Abschnitt 5.2) noch näher behandelt.

4.2.5 Entitäten

Um den fachlichen Ausschnitt des Rated Tags-Ansatzes zu erläutern, werden nachfolgend die relevanten Entitäten und deren Beziehungen zueinander erläutert. Einen Überblick hierzu liefert Abbildung 4.7 in Form eines Entity-Relationship (ER)-Diagramms (in Krähfuß-Notation).

User

Die User-Entität bildet den Endanwender des Systems ab. Ein Anwender hat die Möglichkeit mehrere Rezensionen (Review-Entitäten) zu erstellen. Zudem wird die Erstellung von Benutzer-generierten Rated Tags erlaubt.

Resource

Unter einer Ressource (Resource-Entität) sind hier alle Objekte zu verstehen, welche mit Rated Tags versehen werden können. Hierzu gehören beispielsweise Produkte oder Dienstleistungen. Mögliche Attribute einer Ressource umfassen beispielsweise den Namen oder eine ausführliche Beschreibung. Zu einer Ressource kann es mehrere Rezensionen (Review-Entitäten) geben. Für eine Ressource können ebenfalls mehrere bewertete Tags (RatedTag-Entitäten) vergeben werden. Im Rated Tags-Modell ist eine Ressource immer mindestens einer Kategorie (Category-Entität) zugeordnet.

Review

Eine Review-Entität steht für eine Benutzer-generierte Rezension und ist daher immer genau einer User-Entität zuzuordnen. Die Review-Entität bezieht sich immer auf eine spezielle Ressource (Resource-Entität). Typische Attribute der Review-Entität sind Titel, Text und Gesamtbewertung.

Tag

Die Tag-Entität steht für ein Benutzer-generiertes Schlagwort, welches jedoch im Vergleich zur RatedTag-Entität keine Bewertung umfasst. Der Anbieter verwendet die Tag-Entität für die Vordefinition von Tags (PredefinedTag-Entität).

Um die Qualität des Systems zu verbessern erlaubt das Rated Tags-Modell die explizite Verwendung von Leerzeichen innerhalb eines Tags. Dies ist vor allem dann hilfreich, wenn Aspekte einer Ressource durch mehrere Wörter beschrieben werden, z.B. die englischen Begriffe „operating noise“ oder „wearing comfort“. Durch das Zulassen von Leerzeichen wird sichergestellt, dass durch die Verschlagwortung keine Informationen verloren gehen [57], wie es beispielsweise bei dem Tag „wearingcomfort“ der Fall wäre, da durch die Verknüpfung der beiden Wörter „wearing“ und „comfort“ ohne Leerzeichen die Lesbarkeit des Tags leidet.

Wie Abbildung 4.7 zeigt, ergeben sich keine hierarchischen Verbindungen zwischen einzelnen Tags, d.h., eine hierarchische Strukturierung einzelner Tags ist nicht vorgesehen. Ein Beispiel hierzu wäre die Möglichkeit einem Tag „Bildqualität“ weitere fachlich untergeordnete Tags (wie etwa „Bildschärfe“ oder „Kontrast“) zuzuweisen. Da eine solche Strukturierung im Rahmen eines kollaborativen Tagging Systems, das nicht durch eine zentrale Autorität verwaltet wird, zu komplex erscheint, wird auf eine hierarchische Anordnung verzichtet. Diese würde für das obige Beispiel bedeuten, dass die Tags „Bildqualität“, „Bildschärfe“ und „Kontrast“ sich auf der gleichen Ebene befinden würden.

Rated Tag

Aus Abbildung 4.7 zeigt sich, dass eine RatedTag-Entität eine normale Tag-Entität erweitert. Zu einem bestehenden Tag kann es mehrere RatedTag-Entitäten geben, die sich entweder auf eine Ressource oder eine Rezension beziehen. Aus fachlicher Sicht handelt es sich bei der RatedTag-Entität um die Verknüpfung eines Tags mit einer Bewertung, wobei

die Art der Bewertung nicht vorgeschrieben ist. Allerdings haben die Studienergebnisse aus Abschnitt 3.3.3 gezeigt, dass es sich bei der 5-Punkte Bewertungsskala um die beliebteste Bewertungsform für Gesamtbewertungen in deutschen Onlineshops handelt. Daher erscheint es auch in Hinblick auf eine spätere Benutzbarkeit und Benutzerakzeptanz als sinnvoll, die 5-Punkte Bewertungsskala ebenfalls für die Bewertungskomponente eines Rated Tags heranzuziehen.

Eine spätere Implementierung muss dann über die Darstellungsform der 5-Punkte Bewertungsskala entscheiden. Wie die Studienergebnisse aus Abschnitt 3.3.3 ferner gezeigt haben, wird als Darstellungsform der 5-Punkte Bewertungsskala am häufigsten eine 5-Sterne Notation verwendet. Dieser Gesichtspunkt sollte bei der späteren Implementierung für die Darstellungsform von Rated Tags beachtet werden.

Für die vorgeschlagene 5-Punkte Bewertungsskala spricht also zum einen der Verbreitungsgrad. Zum anderen erlaubt diese Bewertungsform eine relativ detaillierte Bewertung von Aspekten bei einer relativ geringen Komplexität. In einem ähnlichen Ansatz von Vig et al. zur Kombination von Tags und Bewertungen haben sich die Autoren in ihrem Konzept bewusst gegen eine 5-Punkte Bewertungsskala entschieden, da sie davon ausgehen, dass dies die kognitive Komplexität zu stark erhöhen würde [200, S. 325]. Wie jedoch die Benutzerstudie in Kapitel 6 zeigen wird, konnte ein solches Verhalten bei den Anwendern nicht festgestellt werden.

PredefinedTag

Die Entität „PredefinedTag“ dient zur Vordefinition von Tags durch den Betreiber. Daher existiert eine Beziehung zur Tag-Entität. Diese vordefinierten Tags können dann später den Anwendern vorgeschlagen werden. Dadurch können die aus Sicht des Anbieters wichtigsten Tags bereitgestellt werden. Dies kann dazu beitragen, dass die Auswirkungen des im Tagging bestehenden Kaltstartproblems reduziert werden. Ein Betreiber könnte beispielsweise für die Kategorie „Kopfhörer“ (siehe Category-Entität) die Tags „Klangqualität“ und „Tragekomfort“ vordefinieren. Diese Tags könnten dann als Empfehlung für Anwender zur Verfügung gestellt werden.

Category

Wie in Abschnitt 4.2.4 noch näher erläutert wird, soll das Rated Tags-System das Filtern von gleichartigen Ressourcen unterstützen. Ist ein Konsument beispielsweise auf der Suche nach Kopfhörern, so muss es möglich sein, alle für die Ressourcenklasse „Kopfhörer“ vergebenen Rated Tags zu ermitteln. Hierzu ist es folglich erforderlich die Ressourcen entsprechend zu klassifizieren, d.h. ihnen eine Kategorie hinzufügen zu können. Diese Klassifikation ist häufig auch in Onlineshops anzutreffen, wobei hier Produkte und Dienstleistungen oft in hierarchisch aufgebaute Kategorien eingeteilt werden. Dieser Ansatz wird ebenfalls für das Rated Tags-Modell verfolgt, da es eine flexiblere Kategorisierung ermöglicht als die Zuweisung nur einer Kategorie. Wichtig hierfür ist die in Abbildung 4.7 dargestellte Category-Entität. Diese erlaubt die Herstellung einer hierarchischen Struktur, was anhand der Selbstreferenzierung der Entität ersichtlich

wird. Eine Category-Entität kann also mehrere Unterkategorien enthalten, während eine Kategorie zu einer oder keiner Oberkategorie gehört.

Abschließend ist zu erwähnen, dass die in Abbildung 4.7 dargestellten Entitäten nicht mit allen Attributen dargestellt wurden, um die Übersichtlichkeit des ER-Diagramms zu gewährleisten. So werden beispielsweise für die User-Entität keinerlei Attribute (wie etwa Benutzername, Passwort etc.) dargestellt, da diese auf konzeptueller Ebene eine untergeordnete Rolle spielen.

4.2.6 Herausforderungen

Da das Rated Tags-Modell auf den Konzepten des Tagging basiert, ist eine hohe Benutzerbeteiligung eine sehr wichtige Anforderung. Wird ein Produktmerkmal beispielsweise zwar in einer Rezensionen diskutiert, existiert jedoch kein dazugehöriger Tag, so würde eine Abfrage bezüglich dieses Tags diese Rezension nicht zurückliefern. Es ist also eine möglichst einfache Bedienbarkeit notwendig, um die Benutzerbeteiligung und damit die Qualität des Systems zu erhöhen. Eine beispielhafte Erprobung und Evaluierung zur Vergabe und Nutzung von Rated Tags findet sich in der in Kapitel 6 betrachteten Anwenderstudie wieder.

Neben der Benutzerbeteiligung stellt auch die Qualität der erzeugten Tags eine Herausforderung dar. Diesbezüglich lassen sich Tagging-spezifische Probleme identifizieren, die bereits in Abschnitt 3.5.2 diskutiert wurden. Jedoch sind nicht alle Tagging-spezifischen Probleme auf Rated Tags übertragbar. Homonyme stellen für den Rated Tags-Ansatz ein eher geringes Problem dar. Anwender befinden sich bei ihrer Entscheidungsfindung bereits innerhalb einer speziellen Domäne (z.B. innerhalb einer Produktgruppe), wodurch ein geringeres Potenzial hinsichtlich der Mehrdeutigkeiten von Tags besteht. Problematischer sind hingegen Synonyme, die gleiche oder ähnliche Produktmerkmale beschreiben, z.B. die Tags „Klang“ und „Klangqualität“. Zur Adressierung dieses Problems wird in Kapitel 7 ein Modell vorgestellt und evaluiert, um semantisch ähnliche Tags zu gruppieren.

Für die Erstellung von qualitativ hochwertigen Tags sind in der späteren Implementierung des weiteren Regelwerke für die Anwender unverzichtbar, um zu verdeutlichen, welche Form von Tag sich als Rated Tag eignet. Für eine sinnvolle Verarbeitung ist es wichtig, dass Anwender die Neutralität eines Rated Tags sicherstellen. So wäre beispielsweise ein Rated Tag „guter Klang“ weniger hilfreich, da dieser bereits eine Bewertung enthält.

Auch das Vertrauen der Anwender in die erzeugten Rated Tags sowie deren mögliche Manipulation stellen große Herausforderungen dar. Wie ein Artikel der New York Times gezeigt hat, wurden auf den Webseiten des Onlinehändlers Amazon mehrere tausend Rezensionen gelöscht, da diese als manipuliert angesehen wurden [187]. Die Manipulation von Kundenrezensionen beschreiben Hu et al. als die Erzeugung von nicht-authentischen Rezensionen, mit der Absicht den Verkauf eines Produktes anzutreiben [76, S. 674]. Dies kann erreicht werden, indem absichtlich gute oder schlechte Bewertungen vergeben werden. Da Rated Tags ebenfalls eine Bewertungskomponente umfasst, besteht hier also ebenso die Gefahr einer Manipulation. Im Gegensatz zu normalen Rezensionen könnten bei Rated Tags die Manipulation auf einer detaillierteren Ebene stattfinden, z.B. um

die Bewertungen spezieller Produktmerkmale zu manipulieren. Da es sich bei Rated Tags jedoch um ein neuartiges Konzept handelt, kann nicht abgeschätzt werden, ob und wie es für Manipulationszwecke missbraucht wird. Erst ein praktischer Einsatz in einem echten Anwendungsszenario kann zeigen, welche Probleme durch Manipulation oder Spam entstehen.

4.3 Abgrenzung zu verwandten Ansätzen

Nachdem in den vorangegangenen Abschnitten das Konzept von Rated Tags erläutert wurde, soll in diesem Abschnitt dessen Abgrenzung zu bestehenden Arbeiten diskutiert werden. In der nachfolgenden Betrachtung wird Rated Tags sowohl zu bestehenden Lösungsansätzen aus der Praxis als auch zu verwandten wissenschaftlichen Forschungsarbeiten abgegrenzt.

4.3.1 Lösungsansätze aus der Praxis

Wie das Beispielszenario aus Abschnitt 4.1 gezeigt hat, werden die speziellen Konsumentenanforderungen erst bei einer Betrachtung der Rezensionen ersichtlich. Einige Ansätze versuchen jedoch die normalerweise in Rezensionen dargestellten Produktmerkmale einzeln messbar zu machen. Hierzu erlauben Bewertungsportale wie etwa Ciao.de¹ zusätzlich zu einer Gesamtbewertung die Bewertung von detaillierten Produktmerkmalen. Bei einer Digitalkamera könnte zum Beispiel das Merkmal „Bildqualität“ als Bewertungskriterium zur Auswahl stehen. Bei diesem Ansatz werden die Produktmerkmale durch eine zentrale Instanz (z.B. durch das Bewertungsportal oder einen Onlineshop) vordefiniert. Eine grafische Darstellung eines solchen Bewertungsansatzes ist in Abbildung 3.5 auf Seite 44 zu finden.

Ein ähnlicher Ansatz ist die Möglichkeit zur Angabe von Vor- und Nachteilen im Rahmen einer Rezension. Dies soll anderen Konsumenten einen frühzeitigen Überblick über die betrachteten Merkmale und deren Bewertung geben. Ein Beispiel hierzu ist in Abbildung 4.8 dargestellt. Die Abbildung zeigt eine Gesamtbewertung sowie eine kurze Beschreibung von Merkmalen, aufgeteilt nach Pro und Kontra. Da es sich bei den Aufführungen der Pro- und Kontrapunkte ebenfalls um textuelle Bausteine handelt, ist eine automatisierte Verarbeitung sowie eine Wiederverwendung nur schwer möglich. Zudem fehlt dieser binären Bewertungsform (Pro/Kontra) die nötige Granularität. Dies wird in Abbildung 4.8 an der Einordnung des Aspekts „Kamera nur in Ordnung“ sichtbar, was eher einer neutralen Bewertung entspricht.

Ein weiterer Ansatz in einigen aktuellen Onlineshops ist die Möglichkeit Rezensionen anhand eines Suchbegriffes zu filtern. Dabei sollen nur Rezensionen angezeigt werden, welche einen für den Anwender relevanten Suchbegriff enthalten. Ein Anwender könnte beispielsweise nach „Betriebsgeräusch“ suchen, um nur Rezensionen mit diesem Inhalt anzuzeigen. Dieser Ansatz ist allerdings relativ unflexibel, da meist keine komplexen

¹<http://www.ciao.de> (besucht am 03.05.2014)



Abbildung 4.8: Beispiel einer Gesamtbewertung inklusive der Angabe von Vor- und Nachteilen (Quelle: http://www.ciao.de/Samsung_Galaxy_S5_16GB__Test_8994988, besucht am 08.05.2014)

Abfragen erlaubt sind und Produktmerkmale nicht immer in eindeutiger Form beschrieben werden. Würde beispielsweise eine Rezension den Text „*Der Beamer ist kaum hörbar*“ enthalten, so wäre dies eine Beschreibung des Produktmerkmals „Betriebsgeräusch“, welche jedoch nicht anhand dieses Suchbegriffes auffindbar wäre.

4.3.2 Verwandte Forschungsarbeiten

Die an Rated Tags angrenzenden wissenschaftlichen Arbeiten lassen sich in zwei Forschungsbereiche untergliedern, welche nachfolgend näher betrachtet werden. Zum einen werden implizite Verfahren zur automatisierten Analyse von Rezensionen betrachtet, welche versuchen Objektmerkmale und deren Bewertung zu extrahieren. Zum anderen lassen sich Forschungsarbeiten abgrenzen, welche die explizite Verknüpfung von Benutzer-generierten Tags und Bewertungen untersuchen.

Automatisierte Analyse von Rezensionen

Aus wissenschaftlicher Sicht wurden verschiedene Ansätze untersucht, um das in Abschnitt 4.1 genannte Problem zu lösen. Ein Ansatz ist die automatisierte Verarbeitung von Rezensionen mithilfe von Text Mining-Algorithmen. Diese Forschungsansätze versuchen den Text von Rezensionen hinsichtlich der darin enthaltenen Meinungen zu analysieren, um zu bestimmen, ob eine Rezension eine eher positive oder negative Meinung enthält. Der erweiterte Ansatz des Feature-based Opinion Mining versucht zudem die Produktmerkmale aus der Rezension zu ermitteln und die zugehörige Meinung quantitativ zu erfassen.

Im Bereich des Feature-based Opinion Mining existieren mittlerweile sehr viele verschiedene Forschungsarbeiten die sich in ihren Ansätzen, d.h. den Techniken zur Extraktion und Interpretation, unterscheiden. Eine detaillierte Aufführung all dieser Arbeiten wäre an dieser Stelle nicht zielführend. Daher wird für einen Vergleich der bestehenden Ansätze auf diesem Gebiet auf die Arbeiten von Rashid et al. sowie Seerat und Azam verwiesen [162, 174]. Diese beiden Arbeiten liefern einen Überblick über die bestehenden Forschungsarbeiten, die verwendeten Datenquellen sowie über die jeweilige Leistung der Ansätze.

Eine Einschränkung des Feature-based Opinion Mining ist, dass die Bewertungen der Produktmerkmale typischerweise nur als positiv, negativ oder neutral erfasst werden können. Eine Übertragung der Bewertung auf eine Bewertungsskala (z.B. in eine 5-Punkte Bewertungsskala) ist nicht präzise möglich. Diese Einschränkung wird von Pang und Lee als *Rating-Inference Problem* beschrieben [148]. Diese Problematik ist an folgendem Satz erkennbar: „Mir gefällt die Klangqualität des Kopfhörers“. Während ein Opinion Mining Ansatz den vorigen Satz bzw. den dazugehörigen Aspekt „Klangqualität“ beispielsweise als positive Bewertung einstufen könnte, so ist eine exakte Überführung dieser Bewertung auf eine mehrstufige Bewertungsskala (beispielsweise eine 5-Punkte Bewertungsskala) nicht ohne weiteres möglich.

Letztendlich wäre eine automatisierte Überführung in eine Bewertungsskala nur eine Interpretation dessen, was der Rezensent tatsächlich ausdrücken wollte. Eine exakte Bewertung kann jedoch nur explizit durch den Rezensenten erfolgen. Zudem wäre eine automatisierte Interpretation im Hinblick auf das Kundenvertrauen möglicherweise problematisch, denn durch diese Interpretation würde es sich bei den Bewertungen nicht mehr um Benutzer-generierte Daten handeln, sondern um Daten, die durch einen Algorithmus des Anbieters erzeugt wurden. Dies könnte bei Kunden möglicherweise zu Vertrauensproblemen führen.

Die automatisierte Analyse von Rezension mittels Feature-based Opinion Mining bietet noch mehr Herausforderungen [162, 174]. Die Ansätze zur Extraktion und Interpretation von Objektmerkmalen sind oft nur für eine Sprache beschränkt. Diese Ansätze adressieren meistens die englische Sprache und basieren oft auf WordNet¹, einer großen Datenbank im Bereich der englischsprachigen lexikalischen Semantik. Eine Wiederverwendbarkeit für andersartige Sprachen ist daher nicht gegeben, wodurch eine Anpassung erforderlich wird.

Oft wird auch eine manuelle Definition und Interpretation von potenziellen Objektmerkmalen nötig, damit der zugrundeliegende Algorithmus diese erkennt. Ein Beispiel hierzu ist die Ausprägung „sehr stark“. Abhängig vom Kontext, kann dieser Ausdruck positiv oder negativ interpretiert werden. Der Satz „Die Akkuleistung ist sehr stark.“ wäre als positiv einzustufen, wohingegen der Satz „Das Gerät hat eine sehr starke Geräuschentwicklung.“ eher negativ zu bewerten wäre. Gleiches Problem gilt für implizite Bewertungen, z.B. „Der Akku hält 50 Stunden!“. Eine Textanalyse kann hier nicht automatisiert beurteilen, ob diese Information für oder gegen die Akkuleistung spricht.

Abschließend sei erwähnt, dass bestehende Studien aus dem Bereich Opinion Mining zwar Metriken wie Genauigkeit (engl. *precision*) und Trefferquote (engl. *recall*) ermittelt haben, jedoch keine Studie gefunden werden konnte, welche die extrahierten und interpretierten Meinungen als Grundlage für eine Anwenderstudie verwendet hat. Ein direkter Vergleich zu dem hier vorgestellten Rated Tags-Ansatz ist daher nicht möglich. Es wäre jedoch interessant, ob Konsumenten den mittels Opinion Mining generierten Bewertungen vertrauen und wie sich diese Bewertungen auf den Entscheidungsprozess der Konsumenten auswirken.

¹<http://wordnet.princeton.edu> (besucht am 06.06.2014)

Verknüpfung von Benutzer-generierten Tags und Bewertungen

Wie bereits in Abschnitt 3.5.1 erläutert wurde, verwenden Anwender spezielle Tags nicht nur zur Klassifizierung von Inhalten, sondern auch für andere Zwecke. Einer davon ist die Bewertung von Inhalten mithilfe von Tags, wie es diverse Studien belegt haben [6, 53, 129, 175]. Die zentrale Idee des Rated Tags-Ansatzes basiert daher auf der Kombination von Benutzer-generierten Tags und Bewertungen, um die Entscheidungsfindung von Konsumenten zu verbessern. Hinsichtlich der Verknüpfung von Tags und Bewertungen gibt es bereits einige wissenschaftliche Ansätze, welche jedoch andere Forschungsschwerpunkte adressieren. Es folgt eine Erläuterung und Abgrenzung dieser Forschungsarbeiten zum Rated Tags-Ansatz:

Gedikli und Jannach [51]: Diese Arbeit beschäftigt sich in erster Linie mit Recommender Systemen. Hierfür wird ein Konzept vorgeschlagen, welches einen Benutzer-generierten Tag mit einer 5-Sterne Bewertungsskala kombiniert. Der Schwerpunkt der Arbeit liegt nicht in der Bedienbarkeit durch Anwender, sondern in der Evaluierung eines Algorithmus zur Ableitung bzw. Vorhersage von Bewertungen eines Tags. In ihrer Studie wurden Filme untersucht, wobei die Bewertungen der Tags nicht explizit durch Benutzer erzeugt, sondern anhand der Gesamtbewertung interpretiert wurden. Eine Betrachtung von Rezensionen findet hier nicht statt. Während der Fokus der Arbeit in der Vorhersage von Tag-Ratings für Empfehlungen liegt, ist das Ziel von Rated Tags die Nutzung von bestehenden Tag-Ratings zur Verbesserung der Entscheidungsunterstützung.

Lee et al. [116]: Der Forschungsansatz von Lee et al. kombiniert ebenfalls Benutzer-generierte Tags mit einer Bewertungsmöglichkeit. Für die Bewertung kann ein Benutzer hier zwischen den Ausprägungen „negativ“, „neutral“ und „positiv“ wählen. Diese Form der Bewertung wird nachfolgend auch als ternäre Bewertungsskala bezeichnet. Das in der Arbeit beschriebene Tagging System umfasst ebenfalls die Vergabe von normalen Tags, d.h. Tags die zur Inhaltskategorisierung verwendet werden. Diese Tags werden immer mit einer neutralen Bewertung versehen. Zur Präsentation der bewerteten Tags wird eine Tag-Cloud verwendet, welche häufiger verwendete Tags größer darstellt. Mit Hilfe unterschiedlicher Farben werden die Bewertungsarten (positiv, negativ, neutral) voneinander abgegrenzt.

Durch die Bewertungsoption „neutral“ entsteht innerhalb des Tagging Systems jedoch ein Mehrdeutigkeitsproblem. Dies soll an einem Beispiel verdeutlicht werden. Vergibt ein Anwender für eine Kamera den Tag „Zoom“ mit einer neutralen Bewertung, so ließe dies zwei Rückschlüsse zu. Entweder wollte der Anwender damit ausdrücken, dass die Kamera eine Zoom-Funktion besitzt. Dies würde gemäß Sen et al. auf einen faktischen Tag hindeuten [175]. Die zweite Möglichkeit wäre, dass der Anwender den Zoom der Kamera als mittelmäßig oder neutral einstuft. Dies würde auf eine qualitative Bewertung, also auf einen subjektiven Tag, schließen lassen.

Der Ansatz von Lee et al. wurde in einer Anwenderstudie evaluiert. Untersucht wurde dabei das Tagging-Verhalten anhand von vier Experten-generierten Produkt-

rezensionen von ZDNet¹ und vier Webseiten-Artikeln². Der Fokus der Betrachtung lag auf der Tag-Qualität sowie auf der Benutzerakzeptanz. Die Ergebnisse haben gezeigt, dass Anwender gut in der Lage sind, diese Form von bewerteten Tags zu erstellen und zu interpretieren [116]. Ein direkter Bezug zu Rezensionen lag in dieser Studie nicht vor.

Liu et al. [121]: Der Forschungsschwerpunkt von Liu et al. ist die Darstellung von Rezensionen in Onlineshops. Der Ansatz sieht vor, dass Rezensionen hinsichtlich der darin diskutierten Merkmale analysiert werden und dass für diese Merkmale ein Tag mit zugehöriger Bewertung auf einer ternären Skala erzeugt wird. Im Gegensatz zu anderen verwandten Arbeiten, werden bei diesem Ansatz die Tags nicht von den Anwendern selbst, sondern entweder von Produktexperten oder automatisiert durch Algorithmen der Textanalyse erzeugt. Wie bereits im vorigen Abschnitt angesprochen wurde, könnte eine solche automatisierte oder durch Produktexperten generierte Bewertung zu einem Vertrauensproblemen bei Anwendern führen. Abschließend muss erwähnt werden, dass es sich bei dieser Forschungsarbeit lediglich um die Präsentation eines Konzepts handelt. Eine Evaluierung des vorgeschlagenen Ansatzes wurde nicht präsentiert.

Vig et al. [200]: Der Ansatz von Vig et al. ist sehr ähnlich zur Forschungsarbeit von Lee et al. Es wird ebenfalls die Verknüpfung von Benutzer-generierten Tags mit einer ternären Bewertungsskala vorgeschlagen. Wie auch bei Lee et al. erfolgt bei diesem Ansatz keine Trennung zum traditionellen Tagging. Dies führt wie in der Arbeit von Lee et al. zu dem genannten Mehrdeutigkeitsproblem. Ein Bezug zu Rezensionen findet hier ebenfalls nicht statt.

Die Evaluierung des Ansatzes wird durch eine Anwenderstudie vorgenommen, welche die Qualität der Tags, deren Verwendung und vor allem den Einfluss der Benutzerschnittstelle untersucht. Die Evaluierung von Vig et al. basiert auf der Plattform von MovieLens³, welche eine Empfehlungswebseite für Filme darstellt. Die Ergebnisse der Studie waren unter anderem eine verbesserte Benutzerakzeptanz des neuen Tagging Systems im Vergleich zum normalen Tagging, ein Anstieg der vergebenen Tags sowie eine Steigerung der Tag-Qualität. Ferner wurde eine Verringerung an persönlichen Tags festgestellt, was ebenfalls als Qualitätsverbesserung angesehen werden kann, da persönliche Tags laut Sen et al. bei den Anwendern am unbeliebtesten sind [175].

Wie aus den oben dargestellten Forschungsarbeiten ersichtlich ist, existieren bereits Ansätze zur Kombination von Benutzer-generierten Tags und Bewertungen. Allerdings unterscheiden sich deren Anwendungsgebiete und Forschungsschwerpunkte deutlich von Rated Tags. Der Rated Tags-Ansatz besitzt einen starken Bezug zum E-Commerce und im Speziellen eine starke Verknüpfung zu Kundenrezensionen und zielt auf eine Verbesserung

¹<http://www.zdnet.co.kr> (besucht am 14.04.2014)

²<http://www.joins.com> (besucht am 14.04.2014)

³<http://www.movielens.org> (besucht am 14.04.2014)

der Entscheidungsunterstützung ab. Eine Evaluierung des Rated Tags-Modells wurde anhand einer Anwenderstudie durchgeführt, welche in Kapitel 6 näher erläutert wird.

Abschließend präsentiert Tabelle 4.1 eine zusammenfassende Gegenüberstellung der Forschungsarbeiten zur Verknüpfung von Benutzer-generierten Tags und Bewertungen. Wie aus der Tabelle ersichtlich ist, unterscheidet sich der Rated Tags-Ansatz vor allem durch die Betrachtungsweise des Ansatzes hinsichtlich der Entscheidungsfindung. Ferner grenzt sich Rated Tags durch die Benutzer-generierte Definition von bewertbaren Tags im Zusammenhang mit Rezensionen von verwandten Ansätzen ab. Im Gegensatz zu den Arbeiten von Lee et al. und Vig et al. unterscheidet der Rated Tags-Ansatz normale und bewertbare Tags und vermeidet so das Mehrdeutigkeitsproblem, das für die beiden verwandten Ansätzen ([116, 200]) identifiziert wurde.

Tabelle 4.1: Abgrenzung des Rated Tags-Ansatzes zu verwandten Forschungsarbeiten

	Gedikli und Jannach [51]	Lee et al. [116]	Liu et al. [121]	Vig et al. [200]	Rated Tags
Bezeichnung des Forschungsansatzes	Rating items by rating tags	QTag	n/a	Tag Expression	Rated Tags
Forschungsschwerpunkt	Recommender Systeme	Social Tagging	Darstellung von Produktrezensionen	User Interface	Entscheidungsunterstützung
Art der Tag-Erzeugung	Benutzer-generiert	Benutzer-generiert	Durch Experten oder Algorithmen	Benutzer-generiert	Benutzer-generiert
Bewertungsschema	5-Sterne Skala	Ternäre Skala	Ternäre Skala	Ternäre Skala	5-Sterne Skala
Betrachtung von Rezensionen	Nein	Nein	Ja	Nein	Ja
Trennung zu traditionellem Tagging	n/a	Nein	n/a	Nein	Ja
Art der Evaluierung	Evaluierung eines Algorithmus zur Ableitung bzw. Vorhersage von Tag-Ratings	Anwenderstudie bzgl. Tag-Qualität und Verwendung der Tags	n/a	Anwenderstudie bzgl. Tag-Qualität, Verwendung der Tags und Einfluss des User Interfaces	Anwenderstudie hinsichtlich Entscheidungsfindung von Konsumenten (siehe Kapitel 6)
Studienobjekt	Filme	Produkttests und Onlineberichte	n/a	Filme	Elektroprodukte

Kapitel 5

Entwurf und Umsetzung des Rated Tags-Ansatzes

Basierend auf der fachlichen Spezifikation aus Kapitel 4 betrachtet dieses Kapitel den Entwurf und die Umsetzung des Rated Tags-Modells. Zunächst werden die Entwurfsentscheidungen diskutiert. Dabei wird besonders auf die Basisarchitektur sowie auf die Schnittstellenbeschreibung näher eingegangen. Anschließend erfolgt eine Betrachtung des Benutzeroberflächenentwurfs. Darauf aufbauend wird die prototypische Implementierung beschrieben, welche die Grundlage für die in Kapitel 6 beschriebene Anwenderstudie darstellt.

5.1 Systemarchitektur

Der Entwurf einer Softwarearchitektur für das Rated Tags-Modell soll die benötigten Systemkomponenten sowie deren Strukturen und Beziehungen zueinander verdeutlichen. Um die Darstellungskomplexität des Gesamtsystems zu reduzieren und einen besseren Fokus auf dessen einzelne Aspekte zu erhalten, wird beim Entwurf der Softwarearchitektur häufig das Konzept der *Sichten* verwendet [183, S. 77]. Sichten sollen die Architektur eines Systems aus verschiedenen Perspektiven betrachten, um die Vielschichtigkeit der Architektur in angemessener Weise zum Ausdruck zu bringen [183, S. 77]. In der Vergangenheit haben sich verschiedene Modelle zur Definition von Sichten etabliert. Ein bekanntes Sichtenmodell ist das *4+1-Sichten-Softwarearchitekturmodell* von Kruchten [106]. Basierend darauf definiert Starke die folgenden vier Sichten einer Softwarearchitektur [183, S. 79]:

- Kontextsicht
- Verteilungssicht
- Bausteinsicht
- Laufzeitsicht

Die oben genannten Sichten einer Softwarearchitektur werden in den nachfolgenden Abschnitten näher erläutert und für das Rated Tags-Modell beschrieben.

5.1.1 Kontextsicht

Gemäß Starke beschreibt die Kontextsicht der Softwarearchitektur die Einbettung eines Systems in seine Umgebung [183, S. 79]. Im Fall des Rated Tags-Modells wäre die

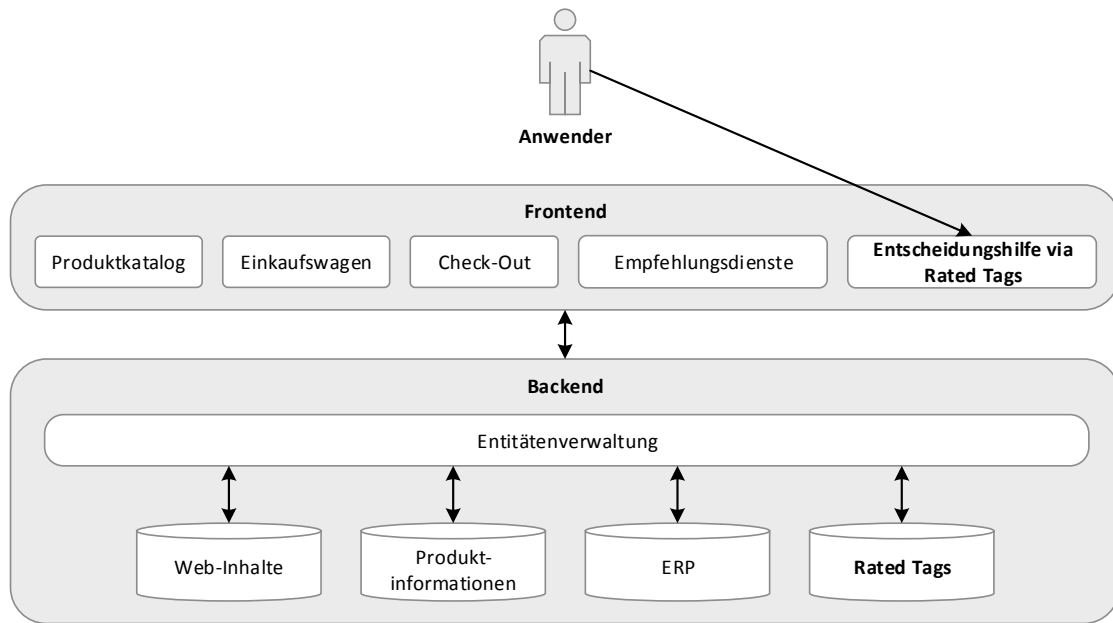


Abbildung 5.1: Einordnung der Rated Tags-Komponenten in die beispielhafte Systemarchitektur eines Onlineshop-Systems aus Abbildung 2.1

entsprechende Umgebung ein Onlinebewertungsportal oder ein Onlineshop. In Bezug auf letztere Variante wird die beispielhafte Systemarchitektur eines Onlineshop-Systems aus Abbildung 2.1 herangezogen. Basierend darauf lassen sich die Komponenten des Rated Tags-Modells in Frontend und Backend unterteilen. Dies ist beispielhaft in Abbildung 5.1 dargestellt. Innerhalb des Frontends sind Rated Tags-Komponenten identifizierbar, welche den Konsumenten des Onlineshops entscheidungsunterstützende Funktionen bereitstellen. Die Grundlage hierfür bilden die Anwendungslogik und die Rated Tags-Entitäten aus dem Backend-Bereich. Die Rated Tags-Entitäten befinden sich auf einer Ebene mit den Produktinformationen und sind mit deren jeweiligen Daten verknüpft.

5.1.2 Verteilungssicht

Basierend auf der vorangegangenen kontextuellen Einordnung des Rated Tags-Modells soll nun die Verteilungssicht betrachtet werden, welche von Starke auch als Infrastruktursicht bezeichnet wird [183, S. 95]. Wie bereits in Kapitel 4 beschrieben wurde, handelt es sich beim Rated Tags-Ansatz um eine Internet-basierte interaktive Entscheidungshilfe, die auf dem Konzept des kollaborativen Tagging basiert. Dies bedeutet, dass eine Vielzahl von verschiedenen Anwendern existiert, welche auf gemeinsame Ressourcen zugreifen und diese auch modifizieren können. Diese Dienstnutzer (*Clients*) greifen dabei auf eine vom Dienstgeber (*Server*) bereitgestellte Funktionalität zurück. Dieses Modell lässt sich allgemein als *Client/Server-Modell* beschreiben, welches gemäß Schill und Springer als traditioneller Ansatz zur Strukturierung von verteilten Systemen angesehen wird [169, S. 14].

Basierend auf dem Client/Server-Modell beschreiben Schill und Springer die weiterentwickelten *objektorientierten*, *komponentenbasierten* und *dienstorientierten* Modelle, welche das Client/Server-Modell um wichtige Aspekte, wie etwa höhere Granularität, Wiederverwendbarkeit und Interoperabilität, erweitern [169, S. 13]. Die objektorientierten und komponentenbasierten Modelle sind typischerweise abhängig von einer bestimmten Plattform. Das dienstorientierte Modell hingegen verfolgt das Ziel einer „Interoperabilität über Plattform- und Unternehmensgrenzen hinweg“ [169, S. 20]. Da die Nutzung des Rated Tags-Systems plattformübergreifend möglich sein soll, z.B. durch Webbrowser, aber auch durch native mobile Anwendungen, wird für den Entwurf des Systems ein dienstorientiertes Modell vorgeschlagen.

Eine Architektur die dem Dienst-orientierten Modell folgt, wird auch als *Service-Oriented Architecture (SOA)* bezeichnet. Das Referenzmodell der *Organization for the Advancement of Structured Information Standards (OASIS)* beschreibt die grundlegenden Konzepte einer SOA. Eine zentrale Bedeutung nimmt diesbezüglich der Begriff „Dienst“ ein, welcher von OASIS wie folgt definiert wird:

„A service is a mechanism to enable access to one or more capabilities, where the access is provided using a prescribed interface and is exercised consistent with constraints and policies as specified by the service description.“ [126, S. 12]

Basierend auf einer dienstorientierten Architektur lassen sich verschiedene Subsysteme ableiten, welche laut Balzert als oberste Abstraktionsebene einer Softwarearchitektur angesehen werden können [10, S. 24]. Die funktionalen Bestandteile der Subsysteme lassen sich generell in die Bereiche von Client und Server einteilen. Die Verarbeitung und Speicherung von systemrelevanten Daten findet üblicherweise im Server statt, während die Präsentation einer Benutzerschnittstelle typischerweise die Aufgabe des Clients ist. In Anlehnung daran wäre eine zweistufige Architektur denkbar. Allerdings hat sich gezeigt, dass eine solche Architektur im Hinblick auf Flexibilität und Austauschbarkeit einige Hürden bereitstellt. Als Erweiterung hierzu haben sich drei- oder sogar mehrstufige Architekturen durchgesetzt [169, S. 23]. Eine dreistufige Architektur (auch *Drei-Schichten-Architektur* genannt) besteht typischerweise aus *Präsentationsschicht*, *Verarbeitungsschicht* und *Persistenzschicht*. Dabei ist die Präsentationsschicht im Client angesiedelt, während sich die Verarbeitungs- und Persistenzschicht typischerweise Server-seitig befindet. Durch die Bildung der einzelnen Schichten ist auf einfache Weise ein Austausch einzelner Schichten möglich. Aus diesem Grund wird für das Rated Tags-System eine Drei-Schichten-Architektur vorgeschlagen.

Eine visuelle Darstellung der für das Rated Tags-System entworfenen Drei-Schichten-Architektur ist Abbildung 5.2 zu entnehmen. Die Abbildung zeigt die oben genannten Schichten, welche nach Client (Service-Nehmer) und Server (Service-Provider) unterteilt sind. Die vom Rated Tags-System verwalteten Entitäten befinden sich im Datenmodell, welches der Persistenzschicht zugeordnet ist. Die Verarbeitung dieser Daten findet durch die Anwendungslogik statt, welche sich in der Verarbeitungsschicht befindet. Diese Schicht ist für die Umsetzung der in Abschnitt 4.2.4 diskutierten Anwendungsfälle verantwortlich. Die Darstellung der Benutzeroberfläche findet letztendlich in der Präsentationsschicht

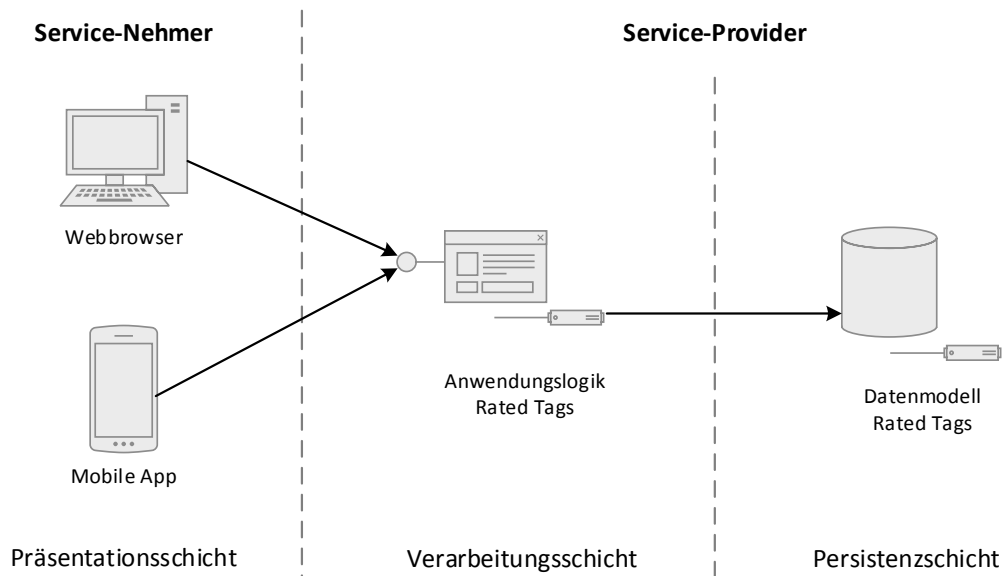


Abbildung 5.2: Drei-Schichten-Architektur des Rated Tags-Systems

statt. In dieser Schicht wird auf die Schnittstellen der Verarbeitungsschicht zugegriffen. Wie aus Abbildung 5.2 zu erkennen ist, sind in der Präsentationsschicht beispielhaft zwei verschiedene Benutzerschnittstellen (Webbrowser und Mobile App) aufgelistet. Dies soll zeigen, dass durch die lose Kopplung der Präsentationsschicht von der Verarbeitungsschicht ein Austausch der verwendeten Benutzerschnittstellen ermöglicht wird.

Ferner ist abschließend anzumerken, dass über die physische Verteilung der Komponenten zu diesem Zeitpunkt nur wenige Angaben gemacht werden. So wird für den Entwurf nur vorgesehen, dass der Service-Nehmer und der Service-Provider physisch voneinander getrennt sind. Die logische Aufteilung der Verarbeitungs- und Persistenzschicht aus Abbildung 5.2 bedeutet nicht zwangsläufig, dass diese beiden Subsysteme später auf physisch getrennte Rechner verteilt werden. Dies obliegt der späteren Implementierungsstrategie.

5.1.3 Bausteinsicht

Die Sichten der vorangegangenen Abschnitte haben die Architektur auf einer hohen Abstraktionsstufe beschrieben. Der nachfolgende Abschnitt soll nun die konkreteren Softwarebausteine (auch *Softwarekomponenten* genannt) sowie deren Schnittstellen und Zusammenhänge näher erläutern. Die Aufteilung von Funktionalität in Softwarekomponenten ist ein wichtiger Schritt zur Erreichung von Wiederverwendbarkeit. Szyperski definiert eine Softwarekomponente dabei wie folgt:

„A software component is a unit of composition with contractually specified interfaces and explicit context dependencies only. A software component can be deployed independently and is subject to composition by third parties.“
[190, S. 41]

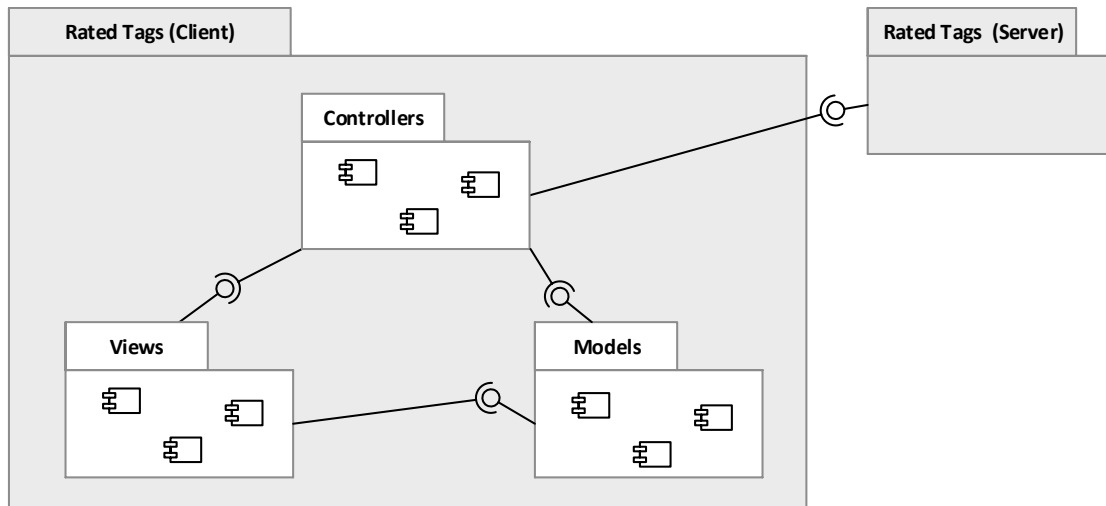


Abbildung 5.3: Abstrakte Darstellung der Client-seitigen Pakete des Rated Tags-Systems basierend auf dem MVC-Entwurfsmuster

Client-seitige Bausteine

Da es sich beim Rated Tags-Modell um eine interaktive Entscheidungshilfe handelt, spielt der Client-seitige Entwurf der Softwarekomponenten eine wichtige Rolle. Für die Konzeption der Anwenderinteraktion hat sich unter anderem das Entwurfsmuster *Model-View-Controller (MVC)* etabliert, welches für den Entwurf des Rated Tags-Systems Client-seitig verwendet werden soll. Bei diesem Entwurfsmuster werden die Darstellung (*View*), die Programmsteuerung (*Controller*) und das Datenmodell (*Model*) in separate Komponenten verlagert, um eine bessere Wartbarkeit und Wiederverwendbarkeit zu erreichen [181, S. 155].

Eine abstrakte Anwendung des MVC-Entwurfsmusters für das Rated Tags-System ist in Abbildung 5.3 dargestellt. Die Abbildung zeigt jeweils den Client und den Server als separate Subsysteme. Das Client-seitige Subsystem zeigt die Pakete *Controllers*, *Views* und *Models* gemäß dem MVC-Entwurfsmuster. Jedes dieser Pakete enthält die konkreten Komponenten für den jeweiligen Anwendungsfall. Typischerweise werden zu einem Anwendungsfall jeweils separate MVC-Komponenten erzeugt, d.h., ein Anwendungsfall umfasst meistens eine eigenständige *View*, ein dediziertes Datenmodell sowie einen *Controller* zur Anwendungssteuerung. Letztlich hängen die jeweiligen Komponenten auch von der verwendeten Oberflächentechnologie ab.

Das Zusammenspiel der einzelnen Komponenten soll an dem nachfolgenden beispielhaften Anwendungsfall illustriert werden. Angenommen ein Anwender selektiert einen Rated Tag, um nur Rezensionen anzuzeigen, die diesen Tag enthalten. Eine Controller-Komponente reagiert dabei auf den Klick des Anwenders und ermittelt den Namen des Tags auf den geklickt wurde. Basierend darauf wird eine Anfrage an den Dienst gestellt, der nur die Rezensionen zurückliefert, die den ausgewählten Tag des jeweiligen Produkts enthalten. Für die weitere Ablaufsteuerung entscheidet die Controller-Komponente

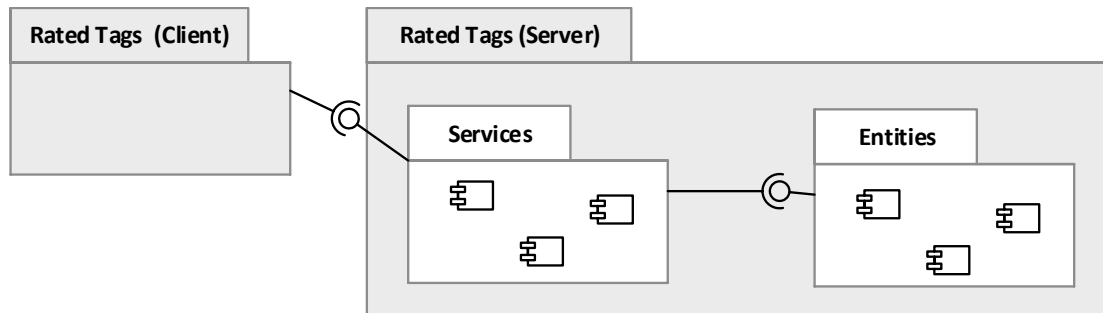


Abbildung 5.4: Abstrakte Darstellung der Server-seitigen Pakete des Rated Tags-Systems

dann basierend auf den empfangenen Daten, welche View-Komponente dem Anwender präsentiert werden soll. Die Controller-Komponente übergibt der ausgewählten View-Komponente dann die empfangenen Daten in Form einer Model-Komponente. Die View-Komponente ist nun in der Lage die Daten entsprechend zu visualisieren und dem Anwender zu präsentieren.

Es ist festzuhalten, dass die Umsetzung der Client-seitigen Komponenten auf Basis des MVC-Musters nur eine Möglichkeit zur Strukturierung der Komponenten ist. Allerdings unterstützt dieses Muster durch die strikte Aufgabenverteilung (engl. *separation of concerns*) die Aufteilung in Komponenten und verbessert somit deren Austauschbarkeit und Wiederverwendbarkeit. Abschließend ist zu erwähnen, dass gemeinsame Funktionalitäten in Server-seitige Komponenten ausgelagert werden sollen, um Redundanzen in den Client-Implementierungen zu vermeiden, d.h., es wird ein sogenannter *Thin Client* angestrebt.

Server-seitige Bausteine

Die Server-seitigen Bausteine des Rated Tags-Systems lassen sich auf unterschiedlichen Abstraktionsebenen beschreiben. Auf hoher Ebene sind die Dienste angesiedelt, welche dem Client die konkreten Funktionalitäten zur Erfüllung der Anwendungsfälle bereitstellen. Diese Dienste wiederum basieren auf verschiedenen Komponenten, welche zur Abarbeitung der Dienstanfragen aufgerufen werden. Weitere Server-seitige Bausteine sind die konkreten Basisentitäten, welche das Datenmodell des Rated Tags-Systems darstellen. Eine Zusammenstellung dieser Komponenten ist in Abbildung 5.4 zu finden. Die Abbildung zeigt auf hoher Abstraktionsebene die beiden Pakete *Services* und *Entities*, welche jeweils die Dienste und Basisentitäten des Systems darstellen.

Wie in Abbildung 5.4 zu erkennen ist, stellen die Dienste (Services-Paket) eine Schnittstelle nach außen zur Verfügung. Als Grundlage für die Kommunikation zwischen den Client- und Server-seitigen Bausteinen ist eine Schnittstellenbeschreibung erforderlich. Gemäß Starke eignet sich hierfür besonders die *Interface Definition Language (IDL)* der *Common Object Request Broker Architecture (CORBA)*, da die IDL eine Technologie-neutrale, Programmiersprachen-unabhängige und deklarative Beschreibung von Schnitt-

stellen zwischen Bausteinen ermöglicht [183, S. 98]. Auf Basis der CORBA IDL zeigt Quelltext 5.1 zunächst die relevanten Server-seitigen Basisentitäten.

Quelltext 5.1: Darstellung der Basisentitäten in CORBA IDL-Notation

```

module Entities
{
    struct Tag
    {
        string Name;
    };

    struct RatedTag
    {
        Tag Tag;
        short Rating;
    };

    struct User
    {
        string Name;
    };

    struct Category
    {
        string Name;
        Category Parent;
    };

    struct Resource
    {
        string Name;
        sequence<Review> Reviews;
        Category Category;
    };

    struct Review
    {
        string Title;
        string Text;
        short OverallRating;
        User User;
        Resource Resource;
        sequence<RatedTag> RatedTags;
    };
};

```

Die in Quelltext 5.1 aufgelisteten Basisentitäten stellen die rein auf Server-seite verwendeten Entitäten dar. Für die Schnittstellenkommunikation zwischen Client und Server sind weitere Entitäten erforderlich, die beiden Parteien bekannt sind. Diese Entitäten werden für den Datenaustausch der Dienstaufrufe verwendet und werden auch des Öfteren als *Data Transfer Objects (DTOs)* bezeichnet [50, S. 401]. Einen Teilausschnitt dieser

Entitäten ist in Quelltext 5.2 dargestellt. Der Quelltext zeigt unter anderem die Entitäten für die aggregierte Darstellung der Rated Tags sowie Entitäten zur Unterstützung eines Filtermechanismus.

Quelltext 5.2: Darstellung der Datentransferentitäten für die Schnittstellenkommunikation in CORBA IDL-Notation

```
module Entities.DataTransfer
{
    struct ResourceDTO
    {
        string Name;
    };

    struct RatedTagAggregate
    {
        string TagName;
        float AverageRating;
    };

    struct RatedTagDTO
    {
        string TagName;
        short Rating;
    };

    struct ReviewDTO
    {
        string Title;
        string Text;
        short OverallRating;
        sequence<RatedTagDTO> RatedTags;
    };

    struct RatedTagsFilter
    {
        string TagName;
        short MinRating;
        short MaxRating;
    };
};
```

Nachdem in Quelltext 5.2 einige relevante Entitäten für die Dienstkommunikation aufgeführt wurden, sollen in Quelltext 5.3 nun die Server-seitigen Schnittstellen zur Kommunikation mit dem Dienst aufgezeigt werden. Die dargestellten Methoden sollen die Kernfunktionalitäten des Rated Tags-Ansatzes verdeutlichen. Hierzu zählt beispielsweise die Reduzierung von Alternativen (`getResources`) oder Rezensionen (`getReviews`). Das Hinzufügen von Rated Tags ist in den Methoden `addTags` und `addReview` angedeutet. Die Methode `getRatedTagsAggregate` ist für die Ermittlung der aggregierten Rated Tags angedacht, um den Anwendern einen schnellen Überblick über die Bewertungen der Rated Tags zu liefern.

Quelltext 5.3: Darstellung der Server-seitigen Schnittstellen in CORBA IDL-Notation

```

module Services
{
    import Entities;
    import Entities.DataTransfer;

    interface ResourceService
    {
        sequence<ResourceDTO> getResources(in Category category,
                                           in sequence<RatedTagsFilter> filters);
    };

    interface ReviewService
    {
        void addReview(in ResourceDTO res, in ReviewDTO review,
                     in User reviewer);
        sequence<ReviewDTO> getReviews(in ResourceDTO r,
                                       in sequence<RatedTagsFilter> filters);
    };

    interface RatedTagsService
    {
        sequence<RatedTagAggregate> getRatedTagsAggregate(in ResourceDTO r);
        sequence<RatedTagDTO> getRatedTags(in ReviewDTO r);
        void addTags(in ReviewDTO r, in User tagger,
                   in sequence<RatedTag> tags);
    };
};

```

Das Ziel dieses Entwurfs ist nicht die komplette Breite an möglicher Funktionalität zu demonstrieren, sondern die Basisfunktionalität zur Unterstützung der in Abschnitt 4.2.4 präsentierten Anwendungsfälle aufzuzeigen. Daher erheben die in den Quelltexten dargestellten Schnittstellen und Operationen keinen Anspruch auf Vollständigkeit. So wird in Quelltext 5.3 beispielsweise nur eine UND-Verknüpfung eines Filters aufgezeigt, während in einer späteren Umsetzung durchaus komplexere Verknüpfungen sinnvoll sein können. Ferner werden auch nur Operationen zur Erzeugung von Tags und Rezensionen aufgezeigt. Eine Verwaltungskomponente zum Editieren und Entfernen von vorhandenen Entitäten findet im Entwurf keine Betrachtung.

Da es sich bei der Definition der Dienstschnittstellen um eine Technologie-unabhängige Darstellung handelt, wird die Art und Weise der Kommunikation der späteren Implementierung überlassen. Denkbare Kommunikationsansätze wären beispielsweise SOAP-basierte Web Services oder Ansätze die dem *Representational State Transfer (REST)*-Paradigma folgen.

5.1.4 Laufzeitsicht

Die vorangegangenen Sichten der Softwarearchitektur haben eher die statischen Aspekte des Modells betrachtet. In diesem Abschnitt soll nun das dynamische Verhalten und das Zusammenwirken ausgewählter Bausteine verdeutlicht werden. Primär werden dabei die

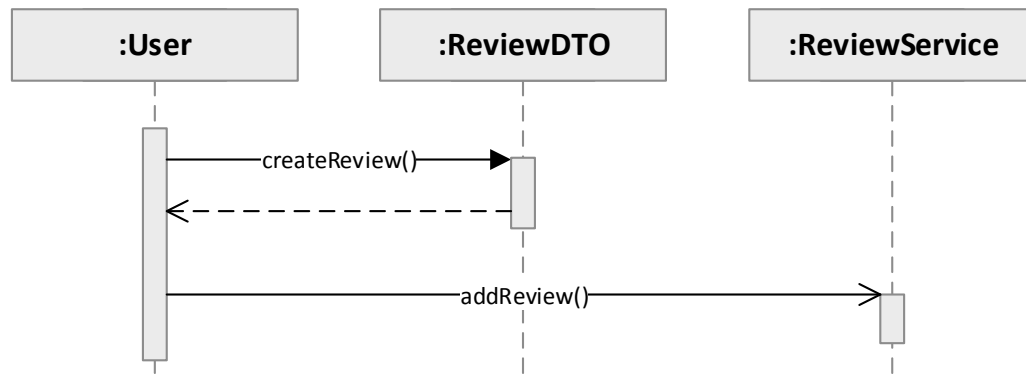


Abbildung 5.5: UML-Sequenzdiagramm für die Erstellung eines Reviews

Interaktionen mit den in Quelltext 5.3 definierten Server-seitigen Schnittstellen näher betrachtet.

Ein zentraler Laufzeitaspekt ist die Erstellung einer neuen Rezension, welche auch Rated Tags enthalten kann. Dies entspricht der Schnittstellenmethode `addReview` aus Quelltext 5.3. Hierbei wird davon ausgegangen, dass ein Anwender zunächst eine spezifische Ressource (z.B. eine Produktseite) aufgerufen hat. Danach können zu dieser Ressource Client-seitig die relevanten Attribute für die Rezension erzeugt werden, z.B. Titel, Gesamtbewertung und Rated Tags. Anhand der erzeugten Objekte folgt dann der Server-seitige Aufruf der Methode `addReview` des Dienstes `ReviewService`. Die Transformation der Client-seitig erzeugten Daten erfolgt dabei in Abhängigkeit zur verwendeten Kommunikationsart, z.B. in Form von XML bei SOAP-basierten Webdiensten. Das Zusammenspiel der oben genannten Komponenten ist in Abbildung 5.5 als UML-Sequenzdiagramm dargestellt.

Eine weitere Ausprägung des `ReviewService`-Dienstes ist die Ermittlung von relevanten Rezensionen mittels der Methode `getReviews`. Der Mehrwert des Rated Tags-Systems besteht unter anderem darin, die Rezensionen basierend auf ihren beigefügten Tags zu filtern. Hierzu stehen dem Anwender zur Laufzeit zwei Möglichkeiten zur Verfügung. Zum einen können innerhalb des Filters nur relevante Tags definiert werden, welche die Gesamtmenge der Rezensionen einschränken. Zum anderen hat der Anwender die Möglichkeit zu den relevanten Tags auch deren Bewertungsgrenzen zu spezifizieren. Hierdurch können Rezensionen ermittelt werden, deren Rated Tags einem bestimmten Grenzwert entsprechen oder diesen über- bzw. unterschreiten. Für diese Art der Filterung kann die Entität `RatedTagsFilter` aus Quelltext 5.2 herangezogen werden. Eine beispielhafte Anwendung eines Filters zur Laufzeit ist in Abbildung 5.6 zu finden. Die Abbildung zeigt die Verwendung der genannten Komponenten, um Rezensionen zu ermitteln, in denen der Tag „Bildqualität“ mit mindestens 4 Sternen und höchstens 5 Sternen bewertet wurde.

Aus dem `RatedTagsService`-Dienst aus Quelltext 5.3 ist vor allem die Methode `getRatedTagsAggregate` interessant, welche für die Aggregation, also für die Bildung der durchschnittlichen Bewertungen über alle Rated Tags, zuständig ist. Hierzu stehen

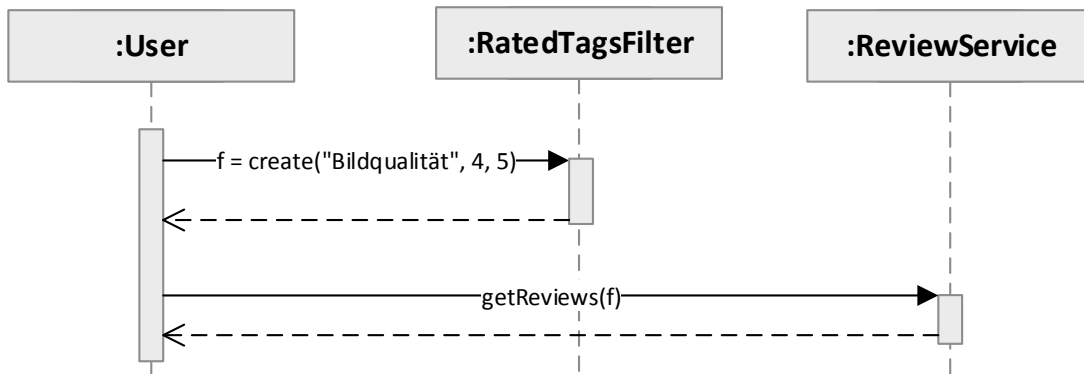


Abbildung 5.6: Beispielhaftes UML-Sequenzdiagramm für eine Filterung von Rezensionen zur Laufzeit

unterschiedliche Alternativen bereit, welche die Performance des Systems zur Laufzeit wesentlich beeinflussen. Die einfachste Variante wäre es, die Aggregation bei jeder Anfrage durchzuführen. Das heißt, wenn sich ein Anwender eine Ressource anzeigen lässt, wird davor eine Operation zur Durchführung der Aggregation angestoßen. Bei dieser Variante wird sichergestellt, dass immer die aktuellsten Bewertungen in die Aggregation mit einfließen. Allerdings hat dieser Ansatz wesentliche negative Auswirkungen auf die Performance, da im Fall von gleichbleibenden Daten eine unnötige Aggregation durchgeführt wird, welche die Reaktionszeit der Anwendung reduziert. Abhilfe zu diesem Problem könnte ein Server-seitiger Caching-Mechanismus liefern, wodurch jedoch die Aktualität des aggregierten Wertes nicht gewährleistet wird.

Eine weitere Möglichkeit wäre eine zyklische Ausführung der Aggregation, z.B. nach einem definierten Zeitintervall. Dies hätte den Vorteil, dass nicht jeder Aufruf eine neue Aggregation auslöst und die Auswirkungen auf die Performance damit besser planbar werden. Ein Nachteil dieses Ansatzes ist ebenfalls eine mögliche unnötige Aggregation, falls sich Rated Tags innerhalb eines Intervalls nicht verändert haben. Ferner ist eine Anpassung der Datenstruktur notwendig, um die zyklisch berechneten Daten zu verwalten. Hinzu kommt, dass bei diesem Ansatz die Möglichkeit besteht, dass Anwender für die Dauer eines Intervalls veraltete Aggregationswerte erhalten. In Abhängigkeit von der gewählten Dauer des Intervalls wäre dies womöglich vertretbar.

Als weitere Alternative ist die Durchführung der Aggregation bei jeder Veränderung eines Rated Tags zu nennen. Dieser Ansatz würde die Nachteile des vorigen Ansatzes eliminieren, da hier immer die aktuellen Werte mit in die Aggregation miteinbezogen werden und keine unnötigen Aggregationen durchgeführt werden. Jedoch ist auch bei diesem Ansatz eine Anpassung der Datenstrukturen notwendig, um die aggregierten Werte in einer passenden Entität zu speichern. Dies wäre bei dem erstgenannten Ansatz, also der stetigen Neuberechnung, nicht erforderlich. Die spätere Implementierung muss die Aspekte der genannten Alternativen abwägen und im Hinblick auf die Anforderungen eine passende Variante auswählen.

5.2 Entwurf der Benutzeroberfläche

Da es sich beim Rated Tags-System um eine interaktive Entscheidungsunterstützung handelt, ist der Entwurf der grafischen Benutzeroberfläche von besonderer Bedeutung. Eine hohe Interaktivität zwischen der Entscheidungshilfe und dem Anwender muss stets gewährleistet werden. Dabei spielt für den Entwurf die verwendete Oberflächentechnologie zunächst eine untergeordnete Rolle. Eine typische und naheliegende Umsetzung wäre eine Web-basierte Oberfläche. Dies ist die am meisten verbreitetste Art im Bereich des E-Commerce. Allerdings haben in den vergangenen Jahren auch immer mehr mobile Endgeräte (z.B. Smartphones oder Tablets) Einzug gehalten, die für den Fortschritt des Mobile Commerce verantwortlich sind. Mobile Endgeräte sind zwar oft auch in der Lage Web-Inhalte über einen Webbrowser abzurufen, jedoch hat die Arbeit von Charland und Leroux gezeigt, dass native Anwendungen auf mobilen Endgeräten in einigen Bereichen (z.B. beim Scrollen von Inhalten) noch eine bessere Bedienbarkeit aufweisen als traditionelle Web-Anwendungen auf diesen Geräten [26, S. 53]. Aufgrund der Heterogenität der Zielplattformen werden daher nachfolgend generische Entwurfsprinzipien für die Oberfläche präsentiert, welche in einer späteren Implementierung sowohl im Web als auch im Bereich der mobilen Anwendungen eingesetzt werden können.

Die von Pfeiffer definierten Anforderungen an ein IIMT (siehe Abschnitt 4.2.4) werden nachfolgend auf den Entwurf der Rated Tags-Benutzeroberfläche angewandt. Zur Erfüllung der Anforderung der *Adaptivität* muss die Benutzeroberfläche so gestaltet werden, dass sie an die Erfahrung des jeweiligen Konsumenten anpassbar ist. Dies bedeutet beispielsweise, dass komplexe Filteroperationen mit mehreren Bedingungen standardmäßig nicht angezeigt werden, sondern erst durch erfahrene Benutzer freigeschaltet werden können.

Die Anforderung der *Flexibilität* kann auf unterschiedliche Weise umgesetzt werden. Zum einen muss es für Anwender möglich sein die Interaktion mit der Entscheidungshilfe zu jedem Zeitpunkt zu beenden, d.h., die Aufhebung eines Filters muss für den Anwender auf einfache Weise möglich sein. Zum anderen muss die Oberfläche auf die unterschiedlichen Bedürfnisse der Anwender angepasst werden können. Dies umfasst beispielsweise die Möglichkeit verschiedene Entscheidungsstrategien anwenden zu können. So muss die Oberfläche beispielsweise eine sortierte Darstellung der Ergebnisse nach bestimmten Kriterien erlauben, um die Anwendung einer lexikographischen Entscheidungsheuristik zu ermöglichen.

Eine weitere von Pfeiffer genannte Anforderung an ein IIMT ist ein *geringer Aufwand* für die Anwender des Systems [155, S. 114]. Durch die aggregierte Darstellung der durchschnittlichen Bewertungen der vergebenen Rated Tags soll der kognitive Aufwand reduziert werden. Weiterhin kann der Aufwand bei der Vergabe von Tags reduziert werden, indem bereits bestehende Tags übersichtlich dargestellt werden und der Anwender während der Eingabe durch vorgeschlagene Tags oder durch eine Autovervollständigung unterstützt wird. Neben dem reduzierten Aufwand fördert dieser Ansatz zudem die Wiederverwendung von Tags. Ferner ist auch die Anzahl der Benutzerinteraktionen (z.B. Mausklicks) zu minimieren, um den Interaktionsaufwand für Anwender so gering wie möglich zu halten. Des Weiteren ist diesbezüglich auch die Verwendung von intuitiv

bedienbaren Steuerelementen erforderlich, welche für den Bereich des E-Commerce bereits verbreitet und allgemein anerkannt sind. Wie die Studie aus Abschnitt 3.3 gezeigt hat, wurde in den meisten untersuchten Onlineshops für die Darstellung der Bewertungen eine 5-Sterne Bewertungsskala verwendet. Daher bietet sich diese bewährte Darstellungsform auch für die Präsentation der Rated Tags-Bewertungen an. Dieser Punkt adressiert auch gleichzeitig die Anforderung an die *Konsistenz* der Benutzeroberfläche. Als weiteren Ansatzpunkt zur Erreichung einer konsistenten Benutzeroberfläche nennen Shneiderman und Plaisant unter anderem die Verwendung von einheitlichen Abkürzungen, Formatierungen und Farben [178, S. 76].

Weitere Anforderungen an die Benutzeroberfläche umfassen die *Steuerung* der Interaktion sowie die *Transparenz* des Systems. Im Hinblick auf das Kriterium der Steuerung muss dem Anwender die volle Kontrolle über die Interaktion gewährt werden. Sämtliche Interaktionen gehen daher von dem Anwender aus und werden nicht über System-seitige Mechanismen beeinflusst, wie dies beispielsweise bei Produktempfehlungen durch Recommendation Systeme der Fall ist. Für die Erfüllung der Transparenz des Systems muss für den Anwender unter anderem der Effekt einer Aktion nachvollziehbar sein. Shneiderman und Plaisant beschreiben Systeme deren Aktionen unverzüglich sichtbar und später ebenfalls rückgängig gemacht werden können als *direct-manipulation interfaces* [178, S. 192]. Die Benutzeroberfläche ist folglich so zu konzipieren, dass der Anwender eine unverzügliche Reaktion auf seine Aktion erhält. Am Beispiel der Filterung bedeutet dies, dass als Reaktion auf eine vom Benutzer gestartete Filter-Aktion entweder eine gefilterte Liste oder eine Statusnachricht über den Fortschritt zurückliefert wird. Im Fall einer Web-basierten Anwendung des Rated Tags-Modells wäre dies beispielsweise durch den Einsatz von dynamischen Elementen basierend auf JavaScript bzw. AJAX möglich. Ist ein Filter aktiv, so muss dies dem Anwender deutlich sichtbar gemacht werden und es muss ihm ferner leicht möglich sein, diesen Filter wieder aufzuheben.

Neben den oben genannten Anforderungen spielt auch die Integration von Rated Tags in bestehende Benutzeroberflächen von Bewertungsportalen oder Onlineshops eine wichtige Rolle. Als Beispiel hierzu lässt sich die Integration in die bekannte *Facettensuche* aus Onlineshops nennen. Bei einer Facettensuche (engl. *faceted search*) werden den Anwendern Metadaten in Form von Schlüssel-Wert-Paaren präsentiert, welche zur Verfeinerung einer Suchanfrage verwendet werden [101, S. 477]. Wie in Abbildung 5.7 zu sehen ist, verwendet der Onlinehändler Amazon eine Facettensuche, um einen Filter anhand von objektiven Produktmerkmalen (hier Kontrastverhältnis und Farbe) zu ermöglichen. In Anlehnung daran lassen sich auch die erzeugten Rated Tags in eine Facettensuche integrieren, da es sich bei Rated Tags ebenfalls um Metadaten handelt, welche jedoch in diesem Fall von den Anwendern selbst generiert werden. Im Gegensatz zu den objektiven Merkmalen aus Abbildung 5.7 entsprechen die Rated Tags den subjektiven Merkmalen, wie beispielsweise der empfundenen Klangqualität. Das entsprechende Steuerelement für eine solche Facettensuche müsste dann die Auswahl der Bewertung und eine Bereichsangabe (z.B. anhand von einfachen Operatoren) erlauben. Eine solche Bereichsangabe ermöglicht Anwendern die detailliertere Einschränkung eines Merkmals, z.B. um nur Produkte anzuzeigen, deren Merkmalsbewertungen über bzw. unter einem definierten Grenzwert liegen.

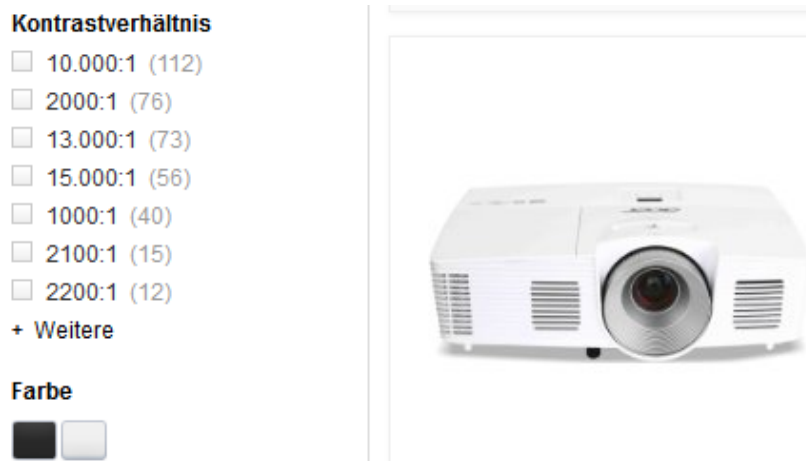


Abbildung 5.7: Beispielhafte Facettensuche auf Amazon.de (Quelle: <http://www.amazon.de>, besucht am 08.07.2014)

Da die Erzeugung der Rated Tags nicht durch eine zentrale Autorität verwaltet wird, kann die Anzahl der unterschiedlichen Tags stark ansteigen. Die Benutzeroberfläche der Facettensuche muss daher so ausgelegt sein, dass eine übersichtliche Darstellung der Tags ermöglicht wird, da die Anzeige aller Tags meist nicht sinnvoll erscheint. Ein möglicher Ansatz wäre die Anzeige der 10 populärsten Tags sowie eine zusätzliche Möglichkeit, um alle weiteren Tags bei Bedarf nachzuladen. Weitere Lösungsansätze die sich mit der optimalen Präsentation von Facetten bei limitierter Bildschirmgröße beschäftigen sind unter anderem der Arbeit von Hearst zu entnehmen [64].

5.3 Prototypische Implementierung

Die prototypische Umsetzung des Rated Tags-Systems verfolgt im Groben zwei Ziele. Zum einen soll mithilfe des Prototyping-Ansatzes die Tauglichkeit des Architekturmodells und der einzelnen Systemkomponenten experimentell nachgewiesen werden [158, S. 27 f.]. Zum anderen dient der entwickelte Prototyp als Basis für eine nachfolgende Evaluierung des Rated Tags-Konzepts. Die allgemeinen Prozessschritte eines Prototyping sind laut Floyd [49, S. 4]:

1. Funktionsauswahl
2. Konstruktion
3. Evaluierung
4. Weitere Verwendung

Der erste Schritt ist zunächst eine Auswahl der Funktionen die für den Prototyp realisiert werden sollen. Dieser Schritt wird in Abschnitt 5.3.1 kurz näher betrachtet. Der nächste Schritt ist die eigentliche Konstruktion des Prototyps, welche in Abschnitt 5.3.2

beschrieben wird. Dieser Schritt umfasst unter anderem die Auswahl der für die Umsetzung verwendeten Technologien. Gemäß dem dritten Prozessschritt aus oben genannter Liste wird der Prototyp dann im Rahmen einer Anwenderstudie hinsichtlich der Entscheidungsfindung von Konsumenten evaluiert (siehe Kapitel 6). Die weitere Verwendung des Prototyps findet dann in Form einer Service-orientierten Umsetzung der Entscheidungshilfe statt. Dieses Vorgehen wird in Kapitel 8 näher betrachtet.

5.3.1 Funktionsauswahl

Für die Auswahl der bereitzustellenden Funktionen war es zunächst notwendig die Art des verwendeten Prototyping zu definieren. Floyd nennt diesbezüglich das *horizontale* und *vertikale* Prototyping, welche sich wie folgt abgrenzen lassen [49, S. 4]: Beim horizontalen Prototyping geht es primär darum eine breite Funktionsabdeckung zu erreichen, wobei einige der verwendeten Funktionen auch simuliert werden können. Im Gegensatz dazu werden beim vertikalen Prototyping ausgewählte Funktionen in ihrer finalen Form implementiert, während die restlichen Funktionen bei der Implementierung noch keine Anwendung finden.

Um den zu erstellenden Prototyp später auch für die Anwenderstudie verwenden zu können, war es notwendig die zentrale Funktionalität in ihrer endgültigen Form umzusetzen, d.h., es wurde ein vertikaler Prototyping-Ansatz gewählt. Für den Prototyp wurden alle für die Anwenderstudie relevanten Anwendungsfälle implementiert, z.B. das Filtern von Rezensionen, die Vergabe von Rated Tags an Rezensionen sowie die aggregierte Darstellung der Rated Tags-Bewertungen. Die zwei Anbieter-spezifischen Anwendungsfälle aus Abschnitt 4.2.4 waren für die Studie nicht relevant und wurden daher für den Prototyp nicht berücksichtigt. Der nachfolgende Abschnitt erläutert die Konstruktion des Prototyps noch näher und präsentiert Screenshots der Benutzeroberfläche.

5.3.2 Konstruktion

Für die Konstruktion des Prototyps kommen grundsätzlich verschiedene Plattformen und Technologien infrage. Wie bereits in Abschnitt 5.2 erwähnt wurde, bieten sich als Plattform hier eine Web-basierte oder eine mobile Anwendung an. Im Hinblick auf die spätere Evaluierung in einer Anwenderstudie wurde als Plattform für den Prototyp allerdings eine Web-basierte Anwendung gewählt, da hier geringere Anforderungen an die späteren Teilnehmer gestellt werden. Bei diesem Ansatz ist lediglich ein Webbrowser erforderlich, wohingegen bei einer Umsetzung in Form einer mobilen Anwendung, jeder Teilnehmer ein mobiles Endgerät, und damit auch ein spezielles Betriebssystem, benötigt hätte.

Aufgrund des Technologie-unabhängigen Entwurfs des Rated Tags-Modells kommen für die Konstruktion der Server-seitigen Logik und Datenhaltung eine Vielzahl von Technologien infrage. Letztendlich fiel die Entscheidung für einen Microsoft-basierten Technologiestack, da hier seitens des Autors bereits ein fundiertes Wissen besteht und somit eine schnellere Entwicklung des Prototyps möglich ist. Konkret bedeutet dies, dass für die Entwicklung der Anwendungslogik die Programmiersprache C# verwendet

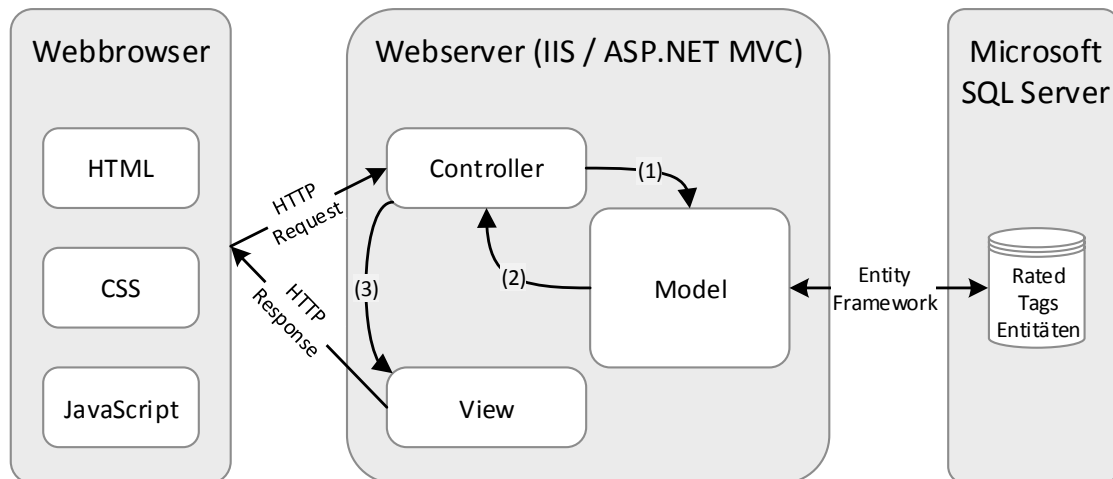


Abbildung 5.8: Technologiestack und logisches 3-Schichtenarchitektur des Prototyps

wurde. Als Web- bzw. Applikationsserver wurde die Dienstplattform *Internet Information Services (IIS)* von Microsoft verwendet. Für eine modulare Entwicklung der Komponenten und die Trennung von Verantwortlichkeiten gemäß dem MVC-Entwurfsmuster wurde das Web-Framework ASP.NET MVC gewählt.

Für die Datenhaltung der Entitäten wurde das relationale Datenbankmanagementsystem *Microsoft SQL Server 2012* in der kostenlosen Express Edition verwendet. Die in Abbildung 4.7 dargestellten Entitäten wurden dementsprechend in eine relationale Datenstruktur überführt. Der Zugriff auf die Entitäten findet über die Anwendungslogik statt. Hier wurde für die Datenzugriffsschicht auf das *Entity Framework* zurückgegriffen. Beim Entity Framework handelt es sich um ein von Microsoft entwickeltes Open Source Framework, das die Abbildung von Objekten auf relationale Datenbankstrukturen (und umgekehrt) erlaubt. Es unterstützt also die Technik der *objektrelationalen Abbildung* (engl. *Object-Relational Mapping, ORM*).

Client-seitig stellt der Webbrowser des jeweiligen Anwenders die Ablaufumgebung für die Nutzung bereit. In diesem Kontext werden daher Technologien wie *HTML*, *Cascading Style Sheets (CSS)* und *JavaScript* verwendet. Aus dieser Ablaufumgebung heraus werden die Anfragen an den Webserver mittels HTTP gestellt und die zurückgelieferten Antworten entsprechend interpretiert. Für eine dynamische Anpassung der Benutzeroberfläche wurde das JavaScript-Framework *jQuery*¹ genutzt. Eine grafische Übersicht der verwendeten Technologien sowie der verwendeten 3-Schichtenarchitektur sind Abbildung 5.8 zu entnehmen.

Neben den Komponenten des Prototyps zeigt Abbildung 5.8 ebenfalls deren Zusammenspiel. Jede eingehende Anfrage wird im Webserver von einer Controller-Komponente entgegengenommen, welche daraufhin die Verarbeitungslogik (Model-Komponente) mit

¹<http://www.jquery.com> (besucht am 02.06.2014)

entsprechenden Parametern versorgt (siehe Schritt (1) aus Abbildung 5.8). Im Fall einer Datenverarbeitung bedient sich die Model-Komponente dem Entity Framework, um Entitäten des Rated Tags-Modells aus der Datenbank zu lesen oder diese dort zu verändern. Nach der Verarbeitung gehen die resultierenden Daten an die Controller-Komponente zurück (siehe Schritt (2)). Diese bestimmt dann die Oberflächendarstellung und ruft die entsprechende View-Komponente mit den Ergebnisdaten auf (siehe Schritt (3)). In der View-Komponente werden die Daten dann entsprechend visuell aufbereitet und im HTML-Format an den Webbrowser zurückgesendet. Im Fall von Anfragen mithilfe von AJAX wird Schritt (3) meist übergangen und es werden nur die reinen Ergebnisdaten (im JSON-Format) an den Webbrowser zurückgeliefert. Hier kümmert sich dann eine Client-seitige Ablaufsteuerung basierend auf JavaScript um die visuelle Aufbereitung der empfangenen Daten. Dies geschieht in der Regel über eine Veränderung des *Document Object Models (DOMs)* des HTML-Dokuments, indem Knoten hinzugefügt, verändert oder entfernt werden.

Für ein besseres Verständnis des Ablaufs auf Serverseite zeigt Quelltext 5.4 beispielhaft die Anwendung des MVC-Entwurfsmusters. Die in Quelltext 5.4 dargestellte Klasse entspricht einer Controller-Komponente. Diese nimmt die Anfragen des Clients entgegen und überträgt die Aufrufe an die in der Klasse definierten Methoden. Welche Aufrufe an welche Methoden weitergeleitet werden, wird in ASP.NET MVC über Konventionen bestimmt, wobei auch manuelle Konfigurationen möglich sind. Die in Quelltext 5.4 gezeigte Methode wäre beispielsweise über den URL-Pfad `/Product/Detail/2` aufrufbar. Das ASP.NET MVC-Framework sucht anhand des URL-Pfads nach einem Controller mit dem Namen `ProductController` und einer Methode `Detail`. Der letzte Teil des URL-Pfads wird als Parameter interpretiert, welcher dann durch das Framework an die Methode weitergeleitet wird (im obiger Beispiel-URL wäre dies der Wert `2`, welcher für den Parameter `productId` verwendet wird).

Quelltext 5.4: Stark verkürzter Auszug einer Controller-Klasse aus dem Webserver

```
public class ProductController : Controller
{
    private IProductService productService;

    public ActionResult Detail(int productId)
    {
        ProductDetailModel model = productService.GetProductDetail(productId);

        return View(model);
    }
}
```

Wie aus der Methode `Detail` aus Quelltext 5.4 hervorgeht, übernimmt die Controller-Komponente nur die Rolle eines Mediators. Die eingehende Produktnummer wird an eine Service-Komponente (`IProductService`) weitergereicht, welche für die Erzeugung des Model-Objekts (in diesem Fall der Produktdetails) zuständig ist. Die Controller-Komponente bestimmt dann die aufzurufende View-Komponente und übergibt ihr das

zurückgelieferte Model-Objekt (`View(model)`). Diese View-Komponente erzeugt dann anhand der übergebenen Daten das entsprechende HTML-Gerüst, welches dann im letzten Schritt per HTTP zurück an den Webbrowser zur Darstellung der Benutzeroberfläche gesendet wird.

Eine beispielhafte, jedoch stark verkürzt dargestellte Implementierung der oben genannten `IProductService`-Schnittstelle ist in Quelltext 5.5 zu finden. Die Implementierung der Schnittstellenmethode `GetProductDetail` ermittelt zunächst anhand der eindeutigen Produktnummer das jeweilige Produkt und alle damit zusammenhängenden Rezensionen. Dies geschieht mithilfe von zwei weiteren Komponenten, welche den Datenzugriff abstrahieren und somit eine leichte Austauschbarkeit der darunterliegenden Datenstruktur sowie eine einfachere Testbarkeit des Systems ermöglichen. Dieser Ansatz ist auch als *Repository*-Entwurfsmuster bekannt. Basierend auf den zurückgelieferten Entitäten wird dann ein Model-Objekt erzeugt, welches die wesentlichen Daten für die Oberflächendarstellung enthält.

Quelltext 5.5: Stark verkürzter Auszug einer Service-Klasse

```
public class ProductService : IProductService
{
    private IProductRepository productRepository;
    private IReviewRepository reviewRepository;

    public ProductDetailModel GetProductDetail(int productId)
    {
        Product p = productRepository.GetById(productId);
        Review[] reviews = reviewRepository.Get(f => f.ProductId == p.Id);

        ProductDetailModel model = new ProductDetailModel();
        model.Product = p;
        model.ProductReviews = reviews;

        return model;
    }
}
```

Wie sich aus dem oben genannten Ablauf ableiten lässt, werden im Prototyp momentan alle Rezensionen zu einem Produkt auf einmal geladen. Dies ist innerhalb des Prototyps vertretbar. In einer späteren Implementierung sollte dies jedoch so abgeändert werden, dass beispielsweise nur die ersten 10 Rezensionen ausgelesen werden und bei Bedarf weitere Rezensionen nachgeladen werden können. Ein typischer Ansatz wäre die Implementierung einer *Paging*-Funktionalität, bei der die Ergebnisliste auf mehrere Seiten aufgeteilt wird und der Anwender jede dieser Seiten einzeln aufrufen kann.

Für eine bessere Verdeutlichung der implementierten Anwendungsfälle werden nachfolgend einige Screenshots präsentiert, welche die Benutzeroberfläche des Prototyps zeigen. Zunächst zeigt Abbildung 5.9 die traditionellen Elemente für die Erzeugung einer neuen Rezension. Die Besonderheiten des Rated Tags-Ansatzes werden dann in Abbildung 5.10 deutlich. Hier werden die Steuerelemente des Prototyps gezeigt, die eine Erzeugung und Wiederverwendung von Rated Tags erlauben. Im oberen Teil von Abbildung 5.10 sind die

Neue Rezension erstellen

Titel: Toller Kopfhörer

Rezension: Die Klangqualität des Kopfhörers ist wirklich sehr gut und er ist auch komfortabel zu tragen. Allerdings stört mich die Länge des Kabels. Ist für meine Zwecke zu kurz.

Gesamtbewertung: ★★★★★

Abbildung 5.9: Beispielhafter Auszug der traditionellen Elemente einer Rezension

von anderen Anwendern vergebenen Rated Tags in alphabetisch sortierter Reihenfolge dargestellt. Der Anwender kann diese Tags wiederverwenden, indem die jeweils passende Sternbewertung durch einen Mausklick selektiert wird. Im Beispiel aus Abbildung 5.10 wird der Tag „Betriebsgeräusche“ wiederverwendet und mit 4 Sternen bewertet. Im unteren Teil der Abbildung hat der Anwender die Möglichkeit neue Rated Tags hinzuzufügen. Hierzu wurde pro Rated Tag ein einfaches Textfeld sowie eine 5-Sterne Bewertungsskala bereitgestellt.

Eine beispielhafte Darstellung der in Abbildung 5.9 erzeugten Rezension ist Abbildung 5.11 zu entnehmen. Die Abbildung zeigt unter anderem die übersichtliche Darstellung der vergebenen Rated Tags. Eine aggregierte Übersicht aller erzeugten Rated Tags ist in Abbildung 5.12 dargestellt. Die Tags sind absteigend nach ihrer Verwendungshäufigkeit sortiert, damit sich die beliebtesten Tags ohne großen Aufwand erkennen lassen. Zu den Benutzern-generierten Tags werden jeweils die Häufigkeiten sowie die durchschnittlichen Bewertungen angezeigt. Dies soll Anwendern einen guten Überblick über die Bewertungen der einzelnen Merkmale liefern. Abbildung 5.12 zeigt ferner die Möglichkeit Rezensionen anhand eines Rated Tags zu filtern. Durch einen Mausklick auf den jeweiligen Tag werden nur die Rezensionen angezeigt, für welche der gewählte Tag vergeben wurde. Im Beispiel aus Abbildung 5.12 wäre dies der Tag „Schwarzwert“. Anhand eines weiteren Mausklicks hat der Anwender die Möglichkeit den gewählten Filter wieder aufzuheben.

Von Nutzern vergebene Rated Tags:

Durch einen Klick auf einen der Sterne können Sie eine Bewertung zu dem jeweiligen Rated Tag vornehmen.

Bedienung	★★★★★ (Zurücksetzen)
Befestigungsmöglichkeit	★★★★★ (Zurücksetzen)
Benutzeroberfläche	★★★★★ (Zurücksetzen)
Betriebsgeräusche	★★★★★ (Zurücksetzen)
Betriebskosten	★★★★★ (Zurücksetzen)

Neue Rated Tags vergeben:

[Hinzufügen...](#)

TestTag ★★★★★ [Entfernen](#)

Abbildung 5.10: Beispielhafte Zuweisung von Rated Tags an eine Rezension

<p>★★★★★ Toller Kopfhörer</p> <p>Author: Testuser</p> <p>Die Klangqualität des Kopfhörers ist wirklich sehr gut und er ist auch komfortabel zu tragen. Allerdings stört mich die Länge des Kabels. Ist für meine Zwecke zu kurz.</p>	<p>Rated Tags</p> <p>Klangqualität ★★★★★</p> <p>Tragekomfort ★★★★★</p> <p>Kabellänge ★★★★★</p>
---	---

Abbildung 5.11: Beispielhafte Darstellung der in Abbildung 5.9 erzeugten Rezension

Übersicht der vergebenen Rated Tags

Bildqualität (15) ★★★★★ (4,3)	Preis/Leistung (6) ★★★★★ (5,0)
Betriebsgeräusche (8) ★★★★★ (4,4)	Lens-Shift (5) ★★★★★ (1,0)
Preis-Leistungsverhältnis (8) ★★★★★ (4,9)	Fernbedienung (4) ★★★★★ (3,3)
Lüftergeräusch (7) ★★★★★ (4,3)	Schwarzwert (3) ★★★★★ (3,3)

Es werden nur Rezensionen mit dem Rated Tag "**Schwarzwert**" angezeigt ([Filter aufheben](#))

Abbildung 5.12: Filtermöglichkeit von Rezensionen anhand von Rated Tags

Kapitel 6

Anwenderstudie zur Evaluierung des Rated Tags-Ansatzes

Basierend auf der prototypischen Implementierung des Rated Tags-Ansatzes aus Kapitel 5 wurde eine Anwenderstudie durchgeführt, um die Einflüsse des Rated Tags-Systems auf die Entscheidungsfindung von Konsumenten zu ermitteln. In Anlehnung an das in Abschnitt 4.2.1 genannte Ziel soll der Rated Tags-Ansatz den Entscheidungsaufwand reduzieren sowie die Entscheidungsqualität der Konsumenten erhöhen. Die nachfolgenden Abschnitte werden das methodische Vorgehen der Studie und deren Ergebnisse genauer erläutern. Abschließend erfolgen eine Interpretation der Ergebnisse sowie die Darstellung von Limitation und Implikationen der Studie. Die nachfolgenden Ausführungen basieren dabei zu einigen Teilen auf dem Artikel „Supporting customers’ decision making with Rated Tags“ [93].

6.1 Methodisches Vorgehen

Nachfolgend wird das methodische Vorgehen zur Planung und Umsetzung der Anwenderstudie beschrieben. Hierzu wurde zunächst ein abstraktes Forschungsmodell konzipiert, welches in Abbildung 6.1 dargestellt ist. In Anlehnung an diese Abbildung lässt sich das grobe Vorgehen der Anwenderstudie wie folgt beschreiben. Zunächst werden fünf zufällig gewählte, vergleichbare Produkte und deren Rezensionen aus den Webseiten des Onlinehändlers Amazon.de extrahiert. Den extrahierten Rezensionen werden dann Rated Tags zugewiesen. Ferner wird anhand der Rezensionen eines der fünf Produkte als dominierendes Produkt hinsichtlich eines gewählten Produktmerkmals bestimmt. Die Studienteilnehmer bekommen dann die Aufgabe sich für das Produkt zu entscheiden, das ihrer Meinung nach hinsichtlich eines vorgegebenen Produktmerkmals am besten abschneidet. Die Teilnehmer werden dabei in eine Kontrollgruppe und eine Rated Tags-Gruppe unterteilt. Die Kontrollgruppe sieht zu den fünf Produkten jeweils nur die Rezensionen, während der Rated Tags-Gruppe zusätzlich die zugehörigen Rated Tags der Rezensionen angezeigt werden. Über den Vergleich des gewählten Produkts eines Teilnehmers mit dem dominierenden Produkt lässt sich dann ermitteln, welche Gruppe häufiger das dominierende Produkt gewählt hat.

Um das oben genannte grobe Vorgehen besser zu verdeutlichen, werden die einzelnen Komponenten und Phasen des Forschungsmodells in den nachfolgenden Abschnitten noch detaillierter betrachtet.

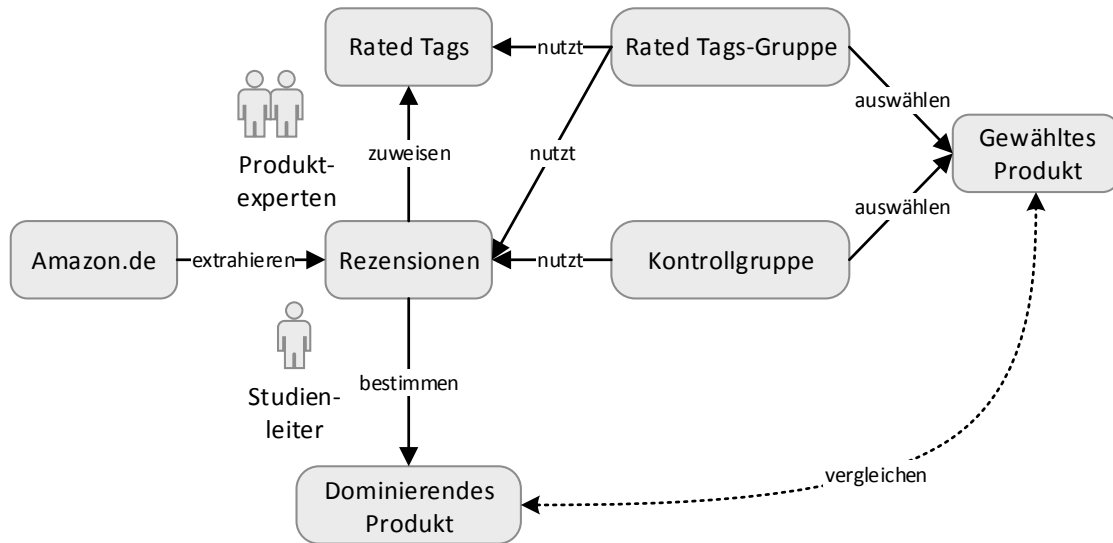


Abbildung 6.1: Forschungsmodell der Anwenderstudie zur Evaluierung des Rated Tags-Ansatzes

6.1.1 Datenbeschaffung und Anpassung des Prototyps

Um zu ermitteln, welchen Einfluss die Verknüpfung von Rated Tags und Rezensionen auf die Entscheidungsfindung von Konsumenten hat, war zunächst die Generierung bzw. Beschaffung von Rezensionen erforderlich. Da es sich bei Rezensionen um echte Benutzermeinungen handelt, kam eine eigene Generierung von Test-Rezensionen nicht infrage. Um den Rated Tags-Ansatz unter echten Bedingungen zu testen, waren also auch echte Kundenrezensionen erforderlich. Für deren Beschaffung wurde auf den Onlineshop Amazon.de zurückgegriffen. Innerhalb der Produktgruppe „Full HD Beamer“ wurden fünf ähnliche Produkte, die jeweils mindestens 20 Rezensionen enthielten, nach dem Zufallsprinzip ausgewählt. Für eine bessere Nachvollziehbarkeit sind die URLs der fünf Produkte in Tabelle 6.1 aufgelistet. Die Tabelle enthält ebenfalls Identifikationsnummern, welche in der Studie anstelle des Produktnamens verwendet wurden. Beispielsweise wurde das Produkt mit Identifikationsnummer 1 in der Studie als „Full HD Beamer #1“ geführt. Die späteren Teilnehmer der Studie kannten weder die echten Produktnamen noch die URLs zu den Produkten.

Um die jeweils ersten 20 Rezensionen eines Produkts zu ermitteln und zu extrahieren, wurde ein *Web Crawler* entwickelt. Ein Web Crawler ist eine Anwendung zur automatisierten Informationsgewinnung [124, S. 265]. Der Web Crawler extrahierte zu jeder Rezension jeweils den Titel, Volltext sowie die der Rezension angehörige Gesamtbewertung auf einer 5-Punkte Skala. Diese Daten wurden dann in die Datenbasis des Prototyps importiert.

Um die spätere Auswahl eines Produkts zu ermöglichen, wurde der in Abschnitt 5.3 beschriebene Prototyp geringfügig erweitert. Auf jeder der fünf Produktseiten wurde ein Button hinzugefügt, damit sich Teilnehmer für das jeweilige Produkt entscheiden können (siehe Abbildung 6.2). Die Selektion eines Produktes wurde erst ermöglicht,

Tabelle 6.1: Identifikationsnummern und URLs zu den in der Studie verwendeten Produkten. Alle URLs wurden zuletzt am 21.11.2013 besucht.

Identifikationsnummer	URL
1	http://www.amazon.de/dp/B002DPMTNQ
2	http://www.amazon.de/dp/B00B2EFZCG
3	http://www.amazon.de/dp/B009SJB6F2
4	http://www.amazon.de/dp/B00B4GJAU
5	http://www.amazon.de/dp/B009JYUORW

Full HD Beamer #1

Sich für dieses Produkt entscheiden

Übersicht der vergebenen Rated Tags

<u>Bildqualität</u> (15)	★★★★★ (4,3)	<u>Preis/Leistung</u> (6)	★★★★★ (5,0)
<u>Betriebsgeräusche</u> (8)	★★★★★ (4,4)	<u>Lens-Shift</u> (5)	★☆☆☆☆ (1,0)
<u>Preis-Leistungsverhältnis</u> (8)	★★★★★ (4,9)	<u>Fernbedienung</u> (4)	★★★★★ (3,3)
<u>Lüftergeräusch</u> (7)	★★★★★ (4,3)	<u>Schwarzwert</u> (3)	★★★★★ (3,3)

Abbildung 6.2: Beispielhafter Ausschnitt einer Produktansicht für Teilnehmer der Rated Tags-Gruppe

nachdem ein Teilnehmer alle Produktseiten mindestens einmal besucht hat. Dadurch sollte vermieden werden, dass ein Teilnehmer ein Produkt auswählt, ohne sich vorher über alle vorhandenen Alternativen informiert zu haben.

6.1.2 Erzeugung von Rated Tags

Nachdem die erforderlichen Kundenrezensionen beschafft wurden, war es notwendig diesen entsprechende Rated Tags zuzuweisen. Eine solche Zuweisung erfolgt normalerweise durch die Rezensenten selbst. Da es sich im Fall der Anwenderstudie jedoch um echte existierende Rezensionen handelt, war eine Zuweisung durch den ursprünglichen Rezensenten nicht möglich. Anstelle dessen wurde die Zuweisung der Rated Tags von fünf ausgewählten Produktexperten vorgenommen. Bei diesen fünf Personen handelt es sich um Mitarbeiter der Hochschule München aus der Fakultät für Informatik und Mathematik. Die ausgewählten Personen besitzen hohes fachliches Wissen über die Produktgruppe der Full HD Beamer.


<p>★★★★★ Top Produkt.... Author: Shawnda Osullivan</p> <p>Ich kann diesen Beamer nur empfehlen, ich habe auf Grund eines Wohnungs Wechsel meinen alten Beamer Benq W6000 verkaufen müssen da die Distanz zu kurz war um eine große Bildprojektion zu bekommenNach 2 Jahren bin ich jetzt auf den Benq w1070 gestoßen um endlich wieder ein 90 Zoll Bild herstellen zu können.... Und dieser Projektor muss sich nicht verstecken</p> <p>+Bildschärfe ist  +Bildeinstellungen sehr umfangreich (für Profis ein Traum) +Lüfter im Gegensatz zu manchen Testberichten sehr leise betrieben im Smart Eco Modus (hier ist mein Laptop im Spielbetrieb ums Vielfache lauter)</p> <p>+/- die Integrierten Lautsprecher sind nicht schlecht jedoch eher für Präsentationen im Büro gedacht als für die Heimkino Fans</p> <p>Also für alle die einen kleinen Raum haben aber nicht auf einen Beamer mit Leinwand verzichten können (so wie ich) kann ich diesen Beamer sehr empfehlen!</p>	<p style="text-align: right;">Rated Tags</p> <p>Betriebsgeräusche ★★★★★★ Bildschärfe ★★★★★★ Einstellungsmöglichkeiten ★★★★★★ Lautsprecher ★★★★★★</p>
---	--

Abbildung 6.3: Beispielhafte Rezension sowie die dafür erstellten Rated Tags

Jeder der fünf Produktexperten hat von den insgesamt 100 Rezensionen jeweils zu 20 Rezensionen Rated Tags erstellt. Die Zuweisung der Rated Tags erfolgte zeitlich versetzt, d.h., ein Produktexperte hat erst mit der Zuweisung begonnen, nachdem der Vorgänger seine Bearbeitung abgeschlossen hatte. Für die Zuweisung wurde eine Eingabemaske entwickelt, die dem aktuellen Produktexperten eine zufällig ausgewählte Rezension aus den noch nicht bearbeiteten Rezensionen präsentiert hat. Ein bereits für eine Rezension vergebener Rated Tag wurde nachfolgend als existierender Tag angezeigt. Jeder Produktexperte hatte somit die Möglichkeit die bestehenden Tags wiederzuverwenden und diese zu bewerten. Eine beispielhafte Darstellung zur Wiederverwendung von Rated Tags ist Abbildung 5.10 zu entnehmen. Das beispielhafte Ergebnis einer mit Rated Tags versehenen Rezension ist in Abbildung 6.3 dargestellt. Die Abbildung zeigt eine Rezension und deren Rated Tags in der Form in der sie später auch den Studienteilnehmern präsentiert wurde.

6.1.3 Bestimmung des dominierenden Produkts

Wie im Forschungsmodell aus Abbildung 6.1 bereits bildlich dargelegt wurde, soll das gewählte Produkt der Teilnehmer mit dem dominierenden Produkt verglichen werden, um die Qualität der Entscheidung zu messen. Hierzu war es zunächst notwendig das dominierende Produkt zu bestimmen. Der Einfachheit halber wurde für diese Bestimmung nur ein Produktmerkmal herangezogen, d.h., ein Produkt das hinsichtlich dieses Produkt-

merkmals am besten abgeschnitten hat, wurde als dominierendes Produkt angesehen. Für die Studie wurde das Produktmerkmal „Betriebsgeräusch“ (nachfolgend auch als M_{BG} abgekürzt) gewählt, um das dominierende Produkt zu bestimmen, d.h., das Produkt mit den geringsten Betriebsgeräuschen sollte das dominierende Produkt darstellen. Eine spezielle Intention hinter der Auswahl dieses Produktmerkmals besteht hierbei nicht. Es hätte ebenso ein anderes, für Beamer typisches Merkmal (z.B. „Bildqualität“) gewählt werden können.

In einer ähnlichen Studie von Häubl und Trifts war das dominierende Produkt führend in allen Produktmerkmalen [62]. Dieses Szenario war allerdings nur möglich, da die Produkte und Merkmalsausprägungen in deren Studie generiert wurden. In der Rated Tags-Studie wurden echte, zufällig ausgewählte Kundenrezensionen extrahiert, weswegen es hier nicht möglich war ein Produkt zu bestimmen, das in allen Merkmalsausprägungen die anderen Produkte dominiert. Im Gegensatz zu der Studie von Häubl und Trifts, welche objektive, generierte Merkmale, wie etwa Gewicht oder Preis, verwendet hat [62, S. 31], basiert die Rated Tags-Studie auf subjektiven Kundenrezensionen. Aus diesem Grund war es zur Bestimmung des dominierenden Produkts notwendig, alle Kundenrezensionen hinsichtlich des Merkmals M_{BG} zu bewerten. Hierzu analysierte der Studienleiter (der Autor der vorliegenden Arbeit) alle Rezensionen und bewertete diese im Hinblick auf das Merkmal M_{BG} auf einer 5-Punkte Skala, wobei eine Ausprägung von 5 Punkten für ein sehr leises Betriebsgeräusch steht.

Die Interpretation von Kundenrezensionen ist allerdings immer subjektiver Natur. Der Satz „Der Beamer ist sehr leise“ könnte von Person A beispielsweise mit 5 Punkten bewertet werden, während Person B hierfür nur 4 Punkte vergibt. Um die Qualität zur Bestimmung des dominierenden Produkts zu erhöhen, wurden daher auch die vergebenen Rated Tags der fünf Produktexperten miteinbezogen. Es hat sich gezeigt, dass die Produktexperten ebenfalls den Tag „Betriebsgeräusch“ identifiziert und verwendet haben. Für jedes Produkt wurden dann alle Rezensionen, zu denen der Tag „Betriebsgeräusch“ vergeben wurde, analysiert, um die durchschnittliche Bewertung zu ermitteln. Abschließend wurden die durchschnittlichen Bewertungen des Studienleiters mit denen der Produktexperten verglichen. Es hat sich zwar gezeigt, dass die Bewertungen hinsichtlich des Tags „Betriebsgeräusch“ etwas voneinander abweichen, jedoch stimmen die Ränge für die fünf untersuchten Produkte überein. Dieser Sachverhalt wird nochmals in Tabelle 6.2 verdeutlicht. Die Tabelle zeigt die durchschnittlichen Bewertungen des Studienleiters und der Produktexperten im Hinblick auf das Merkmal „Betriebsgeräusch“ (M_{BG}). Aus dem Durchschnitt der beiden Bewertungskomponenten (siehe vorletzte Spalte aus Tabelle 6.2) ergibt sich, dass Produkt 1 hinsichtlich des Merkmals M_{BG} alle anderen Produkte dominiert.

Im Hinblick auf das dominierende Produkt ist weiterhin zu erwähnen, dass durch die Festlegung nur eines relevanten Produktmerkmals, den Studienteilnehmern eine lexikographische Entscheidungsheuristik nahegelegt wurde. Bei dieser Entscheidungsstrategie wird das wichtigste Merkmal festgelegt und dann im Hinblick auf die vorhandenen Alternativen untersucht [154, S. 26]. Die ausgewählte Alternative kann dann in Bezug auf dieses Merkmal als dominierend gegenüber den restlichen Alternativen angesehen werden [41, S. 268].

Tabelle 6.2: Durchschnittliche Bewertungen aller Rezensionen auf einer 5-Punkte Skala hinsichtlich des Merkmals „Betriebsgeräusch“ (M_{BG})

Produkt	Ø Bewertung Studienleiter	Ø Bewertung Produktexperten	Bewertung kombiniert	Rang
1	4,67	4,35	4,51	1
2	2,89	2,40	2,65	5
3	3,42	3,25	3,34	3
4	3,00	3,10	3,05	4
5	4,13	3,60	3,87	2

6.1.4 Aufstellung der Hypothesen

Das Ziel der Anwenderstudie ist es, den Einfluss des Rated Tags-Ansatzes auf die Entscheidungsfindung zu überprüfen, d.h., den Mehrwert des Rated Tags-Ansatzes als interaktive Entscheidungshilfe zu ermitteln. Hierzu wurden die folgenden drei Hypothesen aufgestellt, welche im Rahmen der Studie überprüft werden sollen:

Hypothese H1: *Die Verwendung von Rated Tags reduziert bei den Anwendern den Aufwand des Entscheidungsprozesses.*

Es wird die These aufgestellt, dass durch Rated Tags der kognitive Aufwand der Anwender sinkt und dass dadurch eine schnellere Entscheidung möglich ist. Wie bereits durch Lohse und Johnson festgelegt wurde, kann die für eine Entscheidung benötigte Zeit als Kennzahl für den kognitiven Aufwand interpretiert werden [123, S. 89]. Es wird erwartet, dass die Teilnehmer der Rated Tags-Gruppe weniger Zeit für die Entscheidungsfindung benötigen als Teilnehmer der Kontrollgruppe. Es wird davon ausgegangen, dass mindestens zweierlei Faktoren existieren, die zur Reduktion des Aufwandes beitragen. Erstens wird auf jeder Produktseite ein Überblick über alle vergebenen Rated Tags inklusive derer durchschnittlichen Bewertung angezeigt. Dies soll den Teilnehmern einen ersten groben Indikator über die Eignung eines Produkts liefern, wodurch möglicherweise schon erste Alternativen ausgeschlossen werden können. Zweitens ermöglicht die Filtermöglichkeit von Rated Tags, dass nur spezielle Rezensionen angezeigt werden können, welche für die Suche nach dem dominierenden Produkt relevant sind. Dadurch kann die potenzielle Anzahl an zu lesenden Rezensionen reduziert werden, was zu einer zeitlichen Einsparung führen könnte.

Hypothese H2: *Die Verwendung von Rated Tags erhöht die Wahrscheinlichkeit, dass das dominierende Produkt ausgewählt wird.*

In einer perfekten Welt hätten Kunden endlos Zeit und wären daran interessiert alle Kundenrezensionen eines Produkts zu lesen, um eine möglichst durchdachte und optimale Entscheidung zu treffen. Aber die Realität sieht anders aus. Die Zeit ist beschränkt und Entscheidungsträger wollen den Aufwand reduzieren und die

Entscheidungsqualität maximieren [192]. Payne zeigt in seinem Cost-Benefit Framework, dass diese beiden Ziele jedoch oft nicht erreichbar sind, wodurch Entscheider häufig die Entscheidungsqualität gegen einen reduzierten Aufwand eintauschen [152]. Für die Studie wird davon ausgegangen, dass Anwender durch die Filtermöglichkeiten von Rated Tags mehr relevante Rezensionen lesen und so eine höhere Wahrscheinlichkeit haben das dominierende Produkt zu wählen als Teilnehmer der Kontrollgruppe. Dies kann als eine Verbesserung der Entscheidungsqualität angesehen werden.

Hypothese H3: *Durch die Verwendung von Rated Tags erhöht sich die Entscheidungssicherheit der Teilnehmer das richtige Produkt gewählt zu haben.*

Gemäß Yeung und Summerfield umfasst das Konzept der Entscheidungssicherheit die Einschätzung einer Wahrscheinlichkeit, dass die Entscheidung eines Anwenders zu einem gewünschten Resultat führt [217, S. 1311]. Gemäß Häubl und Trifts kann die Entscheidungssicherheit damit als Teil der subjektiven Entscheidungsqualität charakterisiert werden [62]. Da der Rated Tags-Ansatz unter anderem einen aggregierten Überblick der verschiedenen Produkte und deren Merkmale liefert, wird davon ausgegangen, dass dies die Teilnehmer zusätzlich in ihrer Entscheidung bestärkt und sie mit ihrer Entscheidung zufriedener sind als Teilnehmer der Kontrollgruppe. Die Messung der Entscheidungssicherheit erfolgt anhand des Fragebogens über eine 7-Punkte Likert Skala.

Um ermitteln zu können, ob die Manipulation der Versuchsbedingungen, also der unabhängigen Variable, erfolgreich war, wird zusätzlich zu den drei genannten Haupthypothesen die Hypothese aufgestellt, dass es Rated Tags-Teilnehmern im Gegensatz zur Kontrollgruppe leichter fällt die präsentierten Produkte zu vergleichen. Ähnlich zur Studie von Häubl und Trifts [62] wurde dazu eine Frage in den Fragebogen integriert, die Aufschluss darüber geben soll, ob die Manipulation erfolgreich war oder nicht. Dieser Schritt wird im Englischen auch oft als *manipulation check* bezeichnet. Die Teilnehmer aller Gruppen mussten dazu beantworten, wie schwer es ihnen gefallen ist die fünf verschiedenen Produkte miteinander zu vergleichen¹. Als Antwortmöglichkeiten zu dieser Frage wurde eine 7-Punkte Likert Skala bereitgestellt, welche in ihrer Ausprägung von „sehr einfach“ bis „sehr schwer“ reichte.

Abschließend lässt sich festhalten, dass es sich bei den oben genannten Hypothesen jeweils um gerichtete Hypothesen handelt, da behauptet wird, dass der Rated Tags-Ansatz einen positiven Einfluss auf die Entscheidungsfindung von Konsumenten ausübt. Für die spätere Überprüfung der Hypothesen wird ein Signifikanzniveau von $\alpha = 0,05$ festgelegt.

6.1.5 Auswahl geeigneter statistischer Testverfahren

Für die spätere Auswertung der aufgestellten Hypothesen ist es notwendig, vorab die dafür relevanten und passenden statistischen Testverfahren zu identifizieren. Hierzu soll

¹Der genaue Wortlaut ist Frage 7 des Fragebogens zu entnehmen (siehe Tabelle B.1 auf Seite 223).

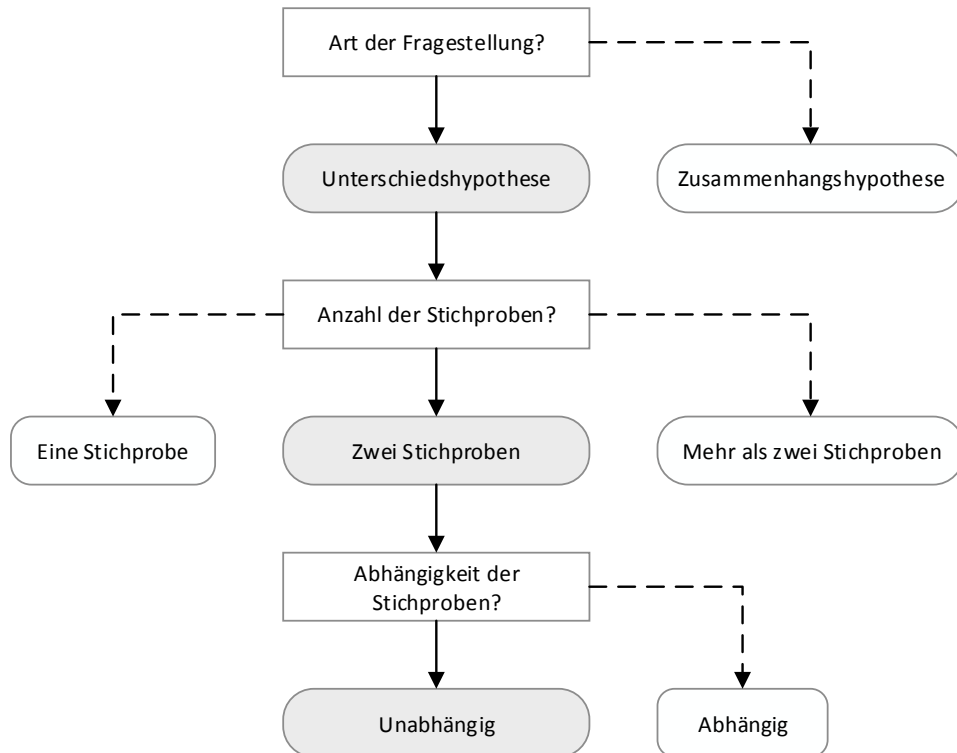


Abbildung 6.4: Einordnung der Studie zur Bestimmung des statistischen Testverfahrens (adaptiert aus [203]). Die schattierten Elemente zeigen die relevanten Ausprägungen der Studie.

als Erstes die Art der Fragestellung, und damit die Art der angewandten Hypothesen, festgestellt werden. Wie bereits erwähnt wurde, soll die Anwenderstudie die Wirksamkeit von Rated Tags untersuchen. Bei den im vorigen Abschnitt genannten Hypothesen handelt es sich daher um *Unterschiedshypothesen*, da sich diese „auf die Wirksamkeit einer Maßnahme oder eines Treatments beziehen“ [20, S. 524].

Für die Wahl des geeigneten Testverfahrens sind ferner die Eigenschaften der verwendeten Stichproben relevant. Da die Aufteilung der Teilnehmer in Rated Tags-Gruppe und Kontrollgruppe erfolgte, handelt es sich folglich um *zwei Stichproben*. Diese können als *unabhängige Stichproben* angesehen werden, da die Beobachtungen der Rated Tags-Gruppe in keinerlei Beziehung zu den Beobachtungen der Kontrollgruppe stehen [21, S. 117]. Die relevanten Ausprägungen der Studie werden nochmals grafisch in Abbildung 6.4 zusammengefasst.

Neben den in Abbildung 6.4 dargestellten Ausprägungen sind ferner die *abhängigen Variablen* sowie deren *Datenniveau* entscheidend, um ein geeignetes Testverfahren auszuwählen. Hier lässt sich gemäß Vorberg und Blankenberger unterscheiden, ob es sich bei der abhängigen Variable um eine *quantitative* oder *qualitative* Messgröße handelt [203]. Bei den Hypothesen H1, H3 sowie bei der Hypothese des Manipulation Checks handelt es sich um quantitativ erfassbare Messgrößen, z.B. die Messung der benötigten

Zeit für Hypothese H1. Bei den quantitativ messbaren Hypothesen geht es in erster Linie um den Vergleich der *Mittelwerte bzw. der zentralen Tendenzen* [203], z.B. wie sich die durchschnittlich benötigte Zeit für die Entscheidung zwischen den beiden Gruppen unterscheidet. Im Gegensatz hierzu handelt es sich bei der abhängigen Variable aus Hypothese H2 um eine qualitative Messgröße.

Bei der abhängigen Variablen aus Hypothese H1 handelt es sich um die Entscheidungsdauer, welche in Millisekunden berechnet wird. Die abhängige Variable ist damit auf einer Verhältnisskala erfassbar und kann somit dem metrischen Datenniveau zugeordnet werden. Als potenziell relevante Tests kommen deshalb der *t-Test* sowie der *Mann-Whitney-U-Test* (Rangsummentest) infrage.

Für Hypothese H2 lässt sich die Entscheidungsqualität als abhängige Variable identifizieren. Die Entscheidungsqualität wird über zwei Kategorien ausgedrückt, nämlich ob das dominierende Produkt ausgewählt wurde oder nicht. Es handelt sich also hier um ein dichotomes Merkmal und somit um ein nominalskaliertes Datenniveau. Die infrage kommenden Tests hierfür sind der *Chi-Quadrat-Test* sowie der *exakte Fisher-Test*. Auf letzteren Test soll dann ausgewichen werden, wenn die Bedingungen für den Chi-Quadrat-Test nicht erfüllt sind, z.B. bei einer sehr geringen Teilnehmeranzahl.

Die Hypothese H3 sowie die Hypothese bezüglich des Manipulation Checks adressieren jeweils die von den Teilnehmern angegebene Entscheidungssicherheit und die Schwierigkeit die Produkte miteinander zu vergleichen. Anhand des Fragebogens wird die Entscheidungssicherheit durch Frage 8 und die Vergleichsschwierigkeit durch Frage 7 abgefragt (siehe Tabelle B.1). Die Antwortmöglichkeiten für beide Fragen liegen in Form einer 7-Punkte Likert Skala vor. Ob Werte einer Likert Skala nicht nur als Ordinal, sondern auch als Intervall interpretiert werden können wird in der Literatur kontrovers diskutiert [84, S. 1217]. Da nicht davon ausgegangen werden kann, dass die Intervalle zwischen den einzelnen Antwortmöglichkeiten gleich groß sind [84, S. 1217], werden Likert Skalen im Rahmen der Studie als Ordinalskalen betrachtet. Daher kommt für die Hypothesen H3 sowie für den Manipulation Check nur der Mann-Whitney-U-Test als Signifikanztest infrage.

Abschließend fasst Tabelle 6.3 die abhängigen Variablen, deren Datenniveaus sowie die potenziellen Signifikanztests der genannten Hypothesen zusammen.

6.1.6 Akquise der Studienteilnehmer

Nach der Vorbereitungsphase und der Anpassung des Prototyps konnte mit der Akquise der Teilnehmer begonnen werden. Zunächst wurden Studierende der Hochschule München¹ aus den Bereichen der Informatik und Wirtschaftsinformatik in Seminaren zur Teilnahme eingeladen. Um eine heterogenere Stichprobe für die Studie zu erzielen, wurden ebenfalls Teilnehmer ohne fundierte Informatikkenntnisse akquiriert. Hierzu wurden Freunde und Bekannte des Studienleiters per E-Mail kontaktiert und um ihre Teilnahme gebeten.

¹<http://cs.hm.edu> (besucht am 17.06.2014)

Tabelle 6.3: Abhängige Variablen, Datenniveaus und potenzielle statistische Tests für die aufgestellten Hypothesen

Hypothese	Abhängige Variable	Daten-niveau	Potenzielle Tests
H1	Entscheidungsdauer (<i>Millisekunden</i>)	Metrisch	<ul style="list-style-type: none"> • t-Test • Mann-Whitney-U-Test
H2	Entscheidungsqualität (<i>dominierendes Produkt gewählt oder nicht</i>)	Nominal	<ul style="list-style-type: none"> • Chi-Quadrat-Test • Exakter Fisher-Test
H3	Entscheidungssicherheit (<i>7-Punkte Likert Skala</i>)	Ordinal	Mann-Whitney-U-Test
Manipulation Check	Vergleichsschwierigkeit (<i>7-Punkte Likert Skala</i>)	Ordinal	Mann-Whitney-U-Test

Der Zeitraum für die Akquise der Teilnehmer sowie für die Durchführung betrug ca. 6 Wochen (14.01.2014 bis 20.02.2014). In dieser Zeit absolvierten insgesamt 34 Teilnehmer die Studie, wobei jeweils 17 Teilnehmer auf die Rated Tags-Gruppe und die Kontrollgruppe entfielen. Es handelte sich bei 18 Teilnehmern um Studierende der Informatik und Wirtschaftsinformatik. Die restlichen 16 Teilnehmer hatten nur geringe Informatikkenntnisse. Von den 34 Teilnehmern waren insgesamt 29 Teilnehmer (ca. 85 %) männlich. Das Alter der Teilnehmer variierte zwischen 18 und 45 Jahren. Eine Verteilung der Altersklassen ist in Abbildung 6.5 dargestellt.

Für die Durchführung der Studie wurde die Domain `ratedtags.com` registriert, über welche die Anwenderstudie abgerufen werden konnte¹. Jedem potenziellen Teilnehmer wurde diese URL mitgeteilt. Um eine gleichmäßige Verteilung der Teilnehmer auf Rated Tags- und Kontrollgruppe zu erreichen, wurden neue Teilnehmer jeweils der Gruppe zugewiesen, die bis dahin weniger Teilnehmer enthielt. War die Teilnehmerzahl identisch, so wurde der Teilnehmer nach dem Zufallsprinzip einer Gruppe zugeordnet. Über einen Identifikator im Query-String der verwendeten URL konnte ermittelt werden, ob es sich um einen Studierenden der (Wirtschafts-)Informatik handelt oder nicht.

Die Aufgabenstellung für Studienteilnehmer war wie folgt. Die Teilnehmer sollten sich in eine Entscheidungssituation versetzen, bei der sie sich aus einer Menge an fünf vergleichbaren Full HD Beamern für eine Alternative entscheiden müssen. Sie sollten

¹Die Domain `ratedtags.com` wurde kurz nach Beendigung der Studie wieder freigegeben und enthält aktuell daher möglicherweise Inhalte, welche nicht mit dieser Arbeit in Verbindung stehen.

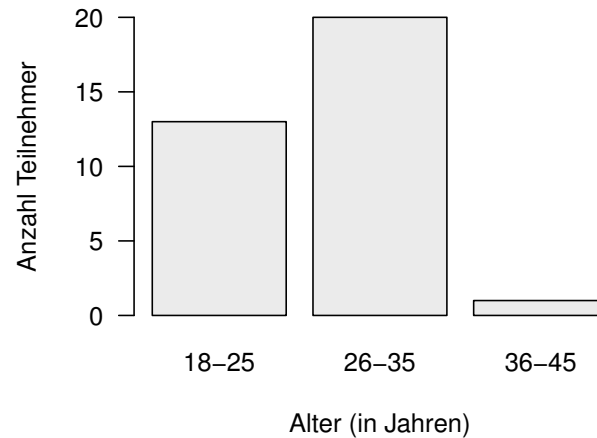


Abbildung 6.5: Verteilung der Altersklassen der Studienteilnehmer (N=34)

dabei davon ausgehen, dass die vorgeschlagenen Alternativen hinsichtlich Bildqualität, Bedienung und weiterer Kriterien bereits begutachtet und für gleichwertig befunden wurden. Lediglich das wichtige Kriterium einer geringen Geräuschentwicklung konnte für die Alternativen noch nicht in Erfahrung gebracht werden. Die Aufgabenstellung ist daher, anhand der vorhandenen Rezensionen, das Produkt mit der geringsten Geräuschentwicklung zu ermitteln. Das Produkt mit dem geringsten Betriebsgeräusch (M_{BG}) wurde bereits in einem vorigen Abschnitt als das dominierende Produkt identifiziert. Ferner wurde innerhalb der Aufgabenstellung angemerkt, dass die Teilnahme an der Studie ohne Pausen durchgeführt werden soll, da die benötigte Zeit für die Studie eine wichtige Rolle spielt.

Zusätzlich zu der oben genannten Aufgabenstellung haben Teilnehmer der Rated Tags-Gruppe eine kurze Erläuterung zum Konzept der Rated Tags erhalten. Es wurde erwähnt, dass es sich bei Rated Tags um Benutzer-definierte Schlagwörter handelt, welche eine Bewertungskomponente enthalten. Den Teilnehmern wurde ferner vermittelt, dass alle Rated Tags von den Autoren der jeweiligen Rezensionen erstellt wurden. Des Weiteren wurde den Teilnehmern mitgeteilt, dass durch einen Klick auf einen Rated Tag nur Rezensionen dargestellt werden, welche diesen Rated Tag enthalten.

Zusätzlich wurde in der Aufgabenstellung erwähnt, dass alle Teilnehmer die sich für das dominierende Produkt entschieden haben, an der Verlosung eines Amazon-Gutscheins im Wert von 20€ teilnehmen. Dies sollte die Motivation der Teilnehmer steigern, sich für das dominierende Produkt zu entscheiden.

Nachdem die komplette Aufgabenstellung von den Teilnehmern gelesen und verstanden wurde, konnte die Studie mit einem Klick gestartet werden. Ab diesem Zeitpunkt wurde die benötigte Zeit aller Teilnehmer gemessen. Sobald sich ein Teilnehmer für ein Produkt entschieden hat, wurde die Zeitmessung beendet. Nach der Auswahl eines Produkts wurde den Teilnehmern ein Fragebogen präsentiert (siehe Tabelle B.1 im Anhang auf Seite 223). Die Teilnehmer waren verpflichtet den Fragebogen komplett auszufüllen. Dies war ebenfalls die Bedingung zur Teilnahme an der oben genannten Verlosung.

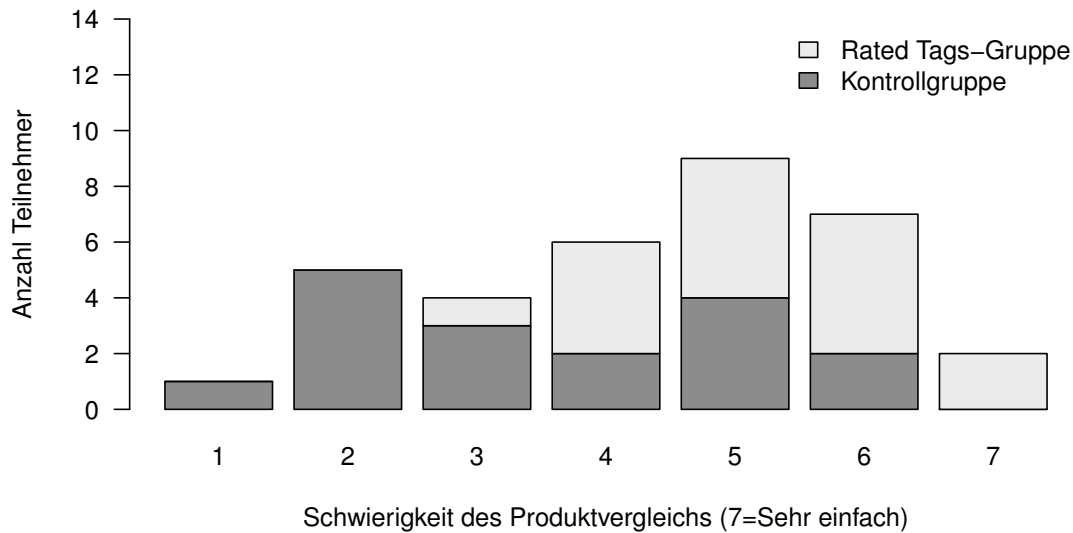


Abbildung 6.6: Schwierigkeit des Produktvergleichs für beide Gruppen basierend auf einer 7-Punkte Likert Skala (N=34)

6.2 Ergebnisse

Nach der Durchführung der Studie gemäß den im vorigen Abschnitt genannten Rahmenbedingungen, sollen nun in diesem Abschnitt die Ergebnisse der Studie präsentiert werden. Für die Auswertung der Daten wurde die freie Programmiersprache R^1 verwendet, welche speziell für die statistische Analyse und Darstellung von Daten entworfen wurde [82, S. 299].

6.2.1 Manipulation Check

Zunächst soll anhand eines Manipulation Checks überprüft werden, ob die Manipulation der Versuchsbedingungen erfolgreich war. Hierzu wird überprüft, ob die Kontrollgruppe mehr Schwierigkeiten bei dem Vergleich der Produkte hatte als die Teilnehmer der Rated Tags-Gruppe. Dies wurde anhand der Frage 7 des Fragebogens (siehe Tabelle B.1 auf Seite 223) ermittelt. Die jeweiligen Antworten der beiden Gruppen sind in Abbildung 6.6 als Säulendiagramm dargestellt. Die Abbildung zeigt, dass ein Vergleich der Produkte für Teilnehmer der Rated Tags-Gruppe einfacher war als für Teilnehmer der Kontrollgruppe. Zur Überprüfung der Signifikanz wurde gemäß Tabelle 6.3 ein einseitiger Mann-Whitney-U-Test durchgeführt. Das Ergebnis des Tests stellte sich als signifikant heraus ($z = 2,87$; $p < 0,05$). Daher wird davon ausgegangen, dass die Manipulation der Versuchsbedingungen erfolgreich war.

¹<http://www.r-project.org> (besucht am 30.06.2014)

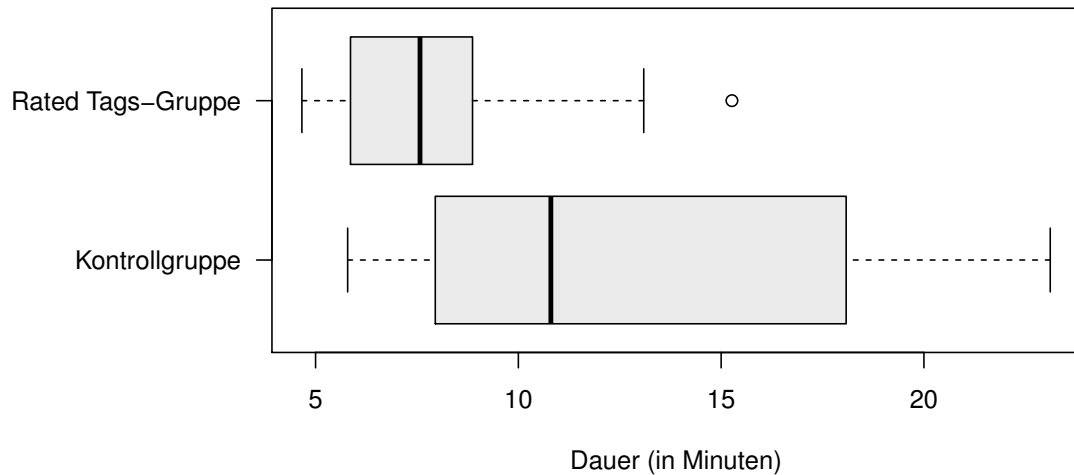


Abbildung 6.7: Boxplot für die Entscheidungsdauer der beiden Gruppen

Tabelle 6.4: Lageparameter für die Entscheidungsdauer der Teilnehmergruppen

	Minimum	1. Quartil	Median	3. Quartil	Maximum
Rated Tags-Gruppe	4,66	5,86	7,58	8,87	15,27
Kontrollgruppe	5,79	7,95	10,80	18,08	23,11

6.2.2 Überprüfung der Hypothesen

In diesem Abschnitt werden die Ergebnisse präsentiert, welche zur Überprüfung der in Abschnitt 6.1 genannten Hypothesen H1 bis H3 beitragen.

Um die Hypothese H1 zu überprüfen musste ermittelt werden, ob der Entscheidungsaufwand in der Rated Tags-Gruppe geringer ist als in der Kontrollgruppe. Hierzu wurde der zeitliche Aufwand, d.h. die Dauer, gemessen, die jeder Teilnehmer vom Beginn der Studie bis hin zur Entscheidung für ein Produkt benötigt hat. Es hat sich gezeigt, dass Teilnehmer der Rated Tags-Gruppe im Mittel 7,98 Minuten für eine Entscheidung benötigt haben, während die durchschnittliche Entscheidungsdauer für Teilnehmer der Kontrollgruppe 12,25 Minuten betrug. Die Verteilung der Entscheidungsdauer für die beiden Gruppen ist in Abbildung 6.7 als Boxplot dargestellt. Die genauen Lageparameter lassen sich in Tabelle 6.4 nachlesen.

Gemäß der Auflistung aus Tabelle 6.3 kommen für Hypothese H1 der t-Test sowie der Mann-Whitney-U-Test infrage. Die Voraussetzungen für einen t-Test verlangen unter anderem, dass das untersuchte Merkmal in den Populationen normalverteilt ist [21, S. 122]. Hiervon kann jedoch für die Entscheidungsdauer nicht ausgegangen werden. Bortz und Schuster beschreiben, dass ein t-Test allerdings auch bei einem entsprechend großen Stichprobenumfang (zentraler Grenzwertsatz) das festgelegte Signifikanzniveau einhält [21, S. 126]. Da der von Bortz und Schuster als Orientierung empfohlene Stichprobenumfang von 30 Beobachtungen pro Stichprobe in der Studie allerdings nicht erfüllt war, wurde auf

Tabelle 6.5: Kontingenztafel mit den Dimensionen „Teilnehmergruppe“ und „Entscheidungsqualität“ (erfolgreich das dominierende Produkt ausgewählt oder nicht)

	Erfolgreich	Nicht erfolgreich	<i>Summe</i>
Rated Tags-Gruppe	14 (R_E)	3 (R_N)	17 (S_R)
Kontrollgruppe	6 (K_E)	11 (K_N)	17 (S_K)
<i>Summe</i>	20 (S_E)	14 (S_N)	34 (N)

eine Verwendung des t-Tests an dieser Stelle verzichtet. Daher wurde zur Überprüfung von Hypothese H1 der voraussetzungsärmere, verteilungsfreie Mann-Whitney-U-Test verwendet, welcher keine Normalverteilungsannahme an die Grundgesamtheit stellt. Ein einseitiger Mann-Whitney-U-Test hat gezeigt, dass es sich bei der kürzeren Entscheidungszeit der Rated Tags-Teilnehmer um ein signifikantes Ergebnis handelt ($z = 2,63$; $p < 0,05$).

Basierend auf dem berechneten z -Wert des Mann-Whitney-U-Tests lässt sich die Effektstärke r berechnen, welche Auskunft über die Größe des Effekts und somit über die praktische Relevanz gibt. Der Veröffentlichung von Effektstärken kommt gemäß Field eine hohe Bedeutung zu, da die Effektstärke ein standardisierter Maß darstellt, welches dadurch unter anderem den Vergleich mit anderen Studien ermöglicht [48, S. 227]. Die allgemeine Berechnung der Effektstärke r ist in Gleichung 6.1 dargestellt, wobei N für die Größe des gesamten Stichprobenumfangs steht. Abgeleitet davon wird in Gleichung 6.2 die Effektstärke r_{saving} berechnet, welche den Effekt der Zeitersparnis repräsentiert. Gemäß der Kategorisierung der Effektstärken (0,1=kleiner Effekt, 0,3=mittelgroßer Effekt, 0,5=großer Effekt) nach Cohen [33, S. 156], kann die berechnete Effektstärke von 0,45 als relativ großer Effekt angesehen werden. Dies deutet auf eine hohe praktische Relevanz der Zeitersparnis durch den Einsatz von Rated Tags hin.

$$r = \frac{z}{\sqrt{N}} \quad (6.1)$$

$$r_{saving} = \frac{2,63}{\sqrt{34}} = 0,45 \quad (6.2)$$

Zur Überprüfung der Hypothese H2 wird zunächst ermittelt, wie viele Teilnehmer sich aus den beiden Gruppen jeweils für das dominierende Produkt entschieden haben. Dies soll Aufschluss über die Unterschiede hinsichtlich der Entscheidungsqualität der beiden Gruppen liefern. Hierzu wurde eine Kontingenztafel mit den Dimensionen „Teilnehmergruppe“ und „Entscheidungsqualität“ erstellt (siehe Tabelle 6.5). Die Kontingenztafel zeigt, dass sich die Teilnehmer der Rated Tags-Gruppe häufiger für das dominierende Produkt entschieden haben als Teilnehmer der Kontrollgruppe.

Um die Signifikanz des Ergebnisses zu überprüfen, wurden in Tabelle 6.3 als mögliche Testverfahren der Chi-Quadrat-Test sowie der exakte Fisher-Test genannt. Zur

Ermittlung der Anwendbarkeit des Chi-Quadrat-Tests müssen zunächst die zu erwartenden Häufigkeiten pro Zelle der Kontingenztafel 6.5 berechnet werden, denn die Autoren Bortz und Schuster empfehlen den Chi-Quadrat-Test nur dann durchzuführen, wenn der Erwartungswert (EW) pro Zelle größer als 5 ist [21, S. 141]. Basierend auf der Kontingenztafel 6.5 werden für diese Berechnung die Zeilensummen S_R und S_K mit den Spaltensummen S_E und S_N multipliziert und jeweils durch die Summe aller Teilnehmer (N) dividiert. Da die Anzahl der Teilnehmer für beide Gruppen, und somit die Zeilensumme, identisch ist ($\frac{N}{2}$), genügt jeweils eine Berechnung für die Erwartungswerte von R_E und K_E (siehe Gleichung 6.3) sowie für R_N und K_N (siehe Gleichung 6.4). Wie die Gleichungen zeigen, ergibt sich für die Zellen R_E und K_E jeweils ein Erwartungswert von 10, während die Zellen R_N und K_N einen Erwartungswert von 7 aufweisen. Somit übersteigen alle Erwartungswerte den geforderten Wert von 5 und erfüllen somit die Anforderung an den Chi-Quadrat-Test. Die nachfolgende Durchführung des Tests hat ergeben, dass die Entscheidungsqualität der Rated Tags-Gruppe signifikant höher ist als die der Kontrollgruppe ($\chi^2 = 5,95$; $p < 0,05$).

$$EW_{R_E} = EW_{K_E} = \frac{\frac{N}{2} \cdot S_E}{N} = \frac{17 \cdot 20}{34} = 10 \quad (6.3)$$

$$EW_{R_N} = EW_{K_N} = \frac{\frac{N}{2} \cdot S_N}{N} = \frac{17 \cdot 14}{34} = 7 \quad (6.4)$$

Um die Stärke dieses Zusammenhangs zu messen, kann unter anderem das Quotenverhältnis (engl. *Odds Ratio*, *OR*) bestimmt werden. Gemäß Field stellt OR eine Effektstärke dar, die den Zusammenhang zwischen Variablen quantifiziert [48, S. 736]. Ein OR lässt sich berechnen, indem die Kreuzprodukte einer Kontingenztafel in Relation zueinander gesetzt werden. Wie aus Gleichung 6.5 zu erkennen ist, ergibt sich basierend auf Kontingenztafel 6.5 für $OR_{dominating}$ der Wert 8,56. Daraus lässt sich ableiten, dass die Chance der Rated Tags-Anwender das dominierende Produkt zu wählen, mehr als 8-mal so hoch war als bei Teilnehmern der Kontrollgruppe.

$$OR_{dominating} = \frac{R_E \cdot K_N}{R_N \cdot K_E} = \frac{14 \cdot 11}{3 \cdot 6} = 8,56 \quad (6.5)$$

Als Letztes soll Hypothese H3 überprüft werden, welche sich auf die Entscheidungssicherheit der Teilnehmer bezieht. Die auf einer 7-Punkte Likert Skala basierenden Ergebnisse der Befragung sind in Abbildung 6.8 als Säulendiagramm dargestellt. Wie das Diagramm veranschaulicht, zeigten sich die Teilnehmer beider Gruppen eher sicher im Hinblick auf die Korrektheit ihrer Entscheidung. Der Modus sowie der Median der Entscheidungssicherheit beider Teilgruppen liegen bei 5. Ein einseitiger Mann-Whitney-U-Test hat ergeben, dass die Entscheidungssicherheit der Rated Tags-Anwender nicht signifikant höher ist als die der Kontrollgruppe ($z = 0,43$; $p > 0,34$). Eine zusammenfassende Darstellung der überprüften Hypothesen sowie derer Ergebnisse ist in Tabelle 6.6 zu finden.

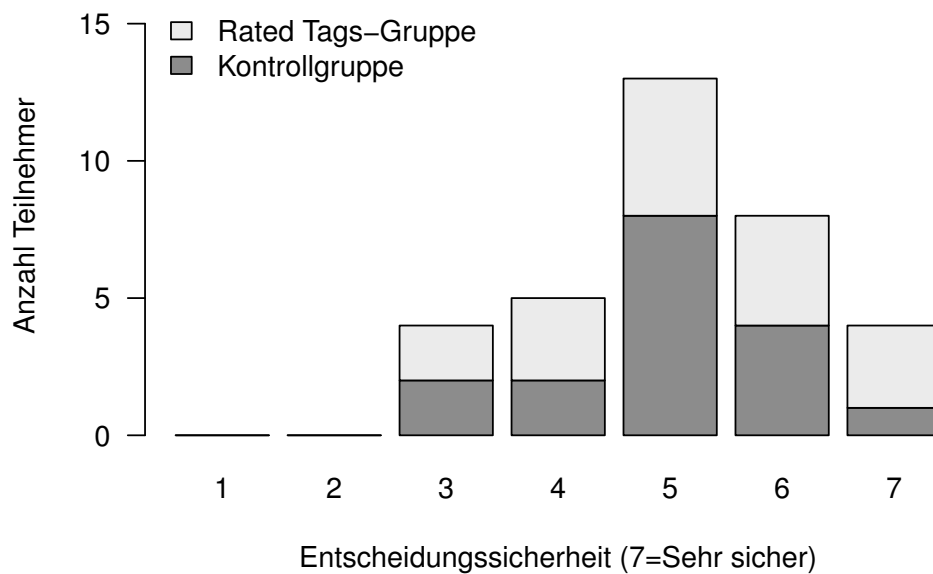


Abbildung 6.8: Entscheidungssicherheit der beiden Gruppen basierend auf einer 7-Punkte Likert Skala (N=34)

Tabelle 6.6: Zusammenfassende Darstellung der gewählten Tests und Ergebnisse für die überprüften Hypothesen

Hypothese	Kurzbeschreibung	Gewählter Test	Ergebnis
Manipulation Check	Rated Tags reduziert die Vergleichsschwierigkeit	Mann-Whitney-U-Test	Nullhypothese wird abgelehnt ($p < 0,05$)
H1	Rated Tags reduziert die Entscheidungsdauer	Mann-Whitney-U-Test	Nullhypothese wird abgelehnt ($p < 0,05$)
H2	Rated Tags erhöht die Entscheidungsqualität	Chi-Quadrat-Test	Nullhypothese wird abgelehnt ($p < 0,05$)
H3	Rated Tags erhöht die Entscheidungssicherheit	Mann-Whitney-U-Test	Nullhypothese wird beibehalten ($p > 0,34$)

6.2.3 Ergebnisse zur Erstellung von Rated Tags

Neben den Erkenntnissen zu den obigen Hypothesen hat die Studie auch einen Einblick in die Benutzer-seitige Erstellung von Rated Tags gegeben. Für die 100 extrahierten Kundenrezensionen wurden insgesamt 382 Rated Tags vergeben, d.h., im Schnitt enthielt jede Rezension fast vier Rated Tags. Die Anzahl an einzigartigen Tags beträgt 57, was auf eine hohe Wiederverwendung bestehender Tags seitens der Anwender schließen lässt. Insgesamt wurden 42 der 57 einzigartigen Tags (ca. 74 %) mehr als einmal verwendet.

Ferner wurde die Qualität der erstellten Tags analysiert. Es hat sich gezeigt, dass von den 57 einzigartigen Tags, 7 Tags (etwa 12 %) als qualitativ minderwertig eingestuft werden können. Die nachfolgende Auflistung zeigt die Gründe hierfür sowie die jeweiligen Tags:

- **Rechtschreibung:** Insgesamt wurde ein erstellter Tag falsch geschrieben: „Zubehörqualität“ anstatt „Zubehörqualität“.
- **Relevanz:** Drei Rated Tags bezogen sich nicht auf das zu bewertende Produkt, sondern auf den Onlinehändler selbst: „Qualität der Lieferung“, „Lieferzeit“ und „Verpackung“.
- **Redundanz:** Drei der vergebenen Tags bezogen sich auf den gleichen Aspekt, wurden jedoch unterschiedlich beschrieben. Hierzu zählt der Aspekt „Zubehör“, welcher durch die Tags „Zubehör“ und „Zubehörqualität“ beschrieben wurde. Der Aspekt „Preis-Leistung“ wurde sogar anhand dreier Tags umschrieben („Preis“, „Preis-Leistungsverhältnis“, „Preis/Leistung“).

Die zehn am häufigsten verwendeten Tags in der Studie können in Tabelle 6.7 nachgelesen werden. Diese können als die am häufigsten diskutierten Aspekte der Kundenrezensionen betrachtet werden.

Neben den reinen Tags ist ferner die Bewertungskomponente der erstellten Rated Tags von Interesse. Hierzu soll zunächst die Beziehung zwischen Rezension und Gesamtbewertung aus traditionellen Bewertungssystemen erläutert werden. Kunden bewerten in Rezensionen typischerweise verschiedene Aspekte und gewichten diese entsprechend [207]. Die resultierende Gesamtbewertung für die Rezension wird durch die einzelnen Aspekte sowie deren Gewichtung entsprechend beeinflusst. Dies hat zur Folge, dass die Gesamtbewertung nicht zwingend der durchschnittlichen Bewertung der Aspekte entspricht, sondern abhängig von den einzelnen Gewichtungen ist. Diese Beziehung soll in Gleichung 6.6 verdeutlicht werden. Die Gleichung zeigt, dass die Gesamtbewertung (B_G) nicht nur abhängig von der Bewertung der einzelnen Aspekte (B_{A_i}) ist, sondern auch die Gewichtung der Aspekte (w_i) berücksichtigt.

$$B_G = \frac{\sum_{i=1}^n w_i B_{A_i}}{\sum_{i=1}^n w_i} \quad (6.6)$$

Da andere Konsumenten die einzelnen Aspekte jedoch anders gewichten können, ist die Aussagekraft einer Gesamtbewertung oft begrenzt. Im Rahmen der Studie wurde geprüft, inwieweit sich die durchschnittliche Bewertung der erstellten Rated Tags (B_R) von der

Tabelle 6.7: Die zehn am häufigsten verwendeten Tags in der Studie (absteigend sortiert nach Häufigkeit)

Tag Name	Häufigkeit
Bildqualität	60
Betriebsgeräusche	41
Lüftergeräusch	26
Preis/Leistung	25
Helligkeit	23
Preis-Leistungsverhältnis	18
Bildschärfe	14
Lens-Shift	12
Anzahl Videoeingänge	9
Regenbogeneffekt	9

Gesamtbewertung (B_G) unterscheidet. Wie Gleichung 6.7 zeigt, bildet B_R den Mittelwert aller Rated Tags Bewertungen, wobei B_{T_i} für die Bewertung eines einzelnen Rated Tags steht. Für die Verdeutlichung des Zusammenhangs zwischen B_G und B_R wird auf Abbildung 6.3 auf Seite 98 verwiesen. Die Abbildung zeigt eine Rezension aus der Studie mit einer Gesamtbewertung von 5 Sternen ($B_G = 5$) sowie vier zugewiesenen Rated Tags. Die durchschnittliche Bewertung aller Rated Tags beträgt hier 4,75 ($B_R = \frac{5+5+5+4}{4} = 4,75$). In diesem Beispiel beträgt die Differenz zwischen B_G und B_R nur 0,25. Basierend auf Gleichung 6.8 wurde jedoch für die durchschnittliche Abweichung (B_A) aller Rezensionen (n) ein Wert von 1,01 ermittelt.

$$B_R = \frac{\sum_{i=1}^n B_{T_i}}{n} \quad (6.7)$$

$$B_A = \frac{\sum_{i=1}^n |B_{G_i} - B_{R_i}|}{n} \quad (6.8)$$

Um die Signifikanz des Ergebnisses zu messen, wurde ein zweiseitiger Mann-Whitney-U-Test durchgeführt. Es hat sich gezeigt, dass sich die Mittelwerte zwischen B_G und B_R in hoch signifikanter Weise unterscheiden ($z = 7,52$; $p < 0,001$). Die Implikationen dieses Ergebnisses werden in Abschnitt 6.3 noch näher diskutiert.

6.2.4 Ergebnisse des Fragebogens

Neben den primären Ergebnissen der Studie hat auch die Auswertung des Fragebogens weitere Erkenntnisse über die Studienteilnehmer geliefert. Alle Teilnehmer antworteten,

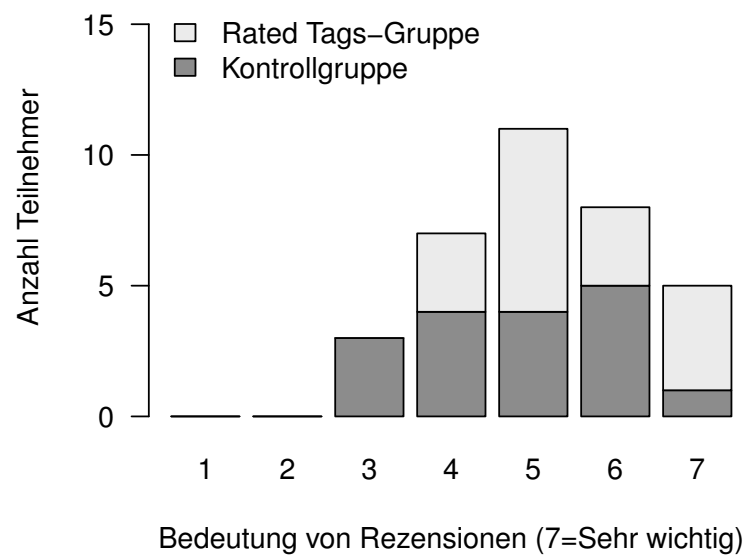


Abbildung 6.9: Antworten der Teilnehmer zur Bedeutung von Kundenrezensionen basierend auf einer 7-Punkte Likert Skala (N=34)

dass sie im Rahmen einer Kaufentscheidung in der Vergangenheit mindestens einmal Kundenrezensionen gelesen haben. Die Bedeutung von Kundenrezensionen wurde dabei von vielen Teilnehmern als hoch bis sehr hoch eingestuft. Die genaue Verteilung der Antworten ist in Abbildung 6.9 als Säulendiagramm dargestellt.

Ferner wurde ermittelt, dass etwa 35 % der Teilnehmer bereits selbst Kundenrezensionen verfasst haben. Die Frage an die Rated Tags-Teilnehmer, ob sie bereit wären Rated Tags selbst zu nutzen, wurde von ca. 82 % der Befragten bejaht. Interessanterweise wurde die Frage insgesamt von allen Teilnehmern bejaht, welche bereits in der Vergangenheit eigene Kundenrezensionen verfasst haben. Die Frage ob sich die Teilnehmer auch eine einzelne Nutzung von Rated Tags, d.h. ohne eine Verbindung zu Rezensionen, vorstellen könnten, wurde nur von einer Person bejaht. Dieses Ergebnis deutet an, dass Bewertungen alleine, d.h. ohne Rezension, für Kunden nicht ausreichen, um eine fundierte Entscheidung zu fällen.

Im Rahmen des Fragebogens wurde weiterhin abgefragt, wie die Teilnehmer ihr Produktwissen hinsichtlich der Produktgruppe „Full HD Beamer“ auf einer 7-Punkte Likert Skala einschätzen würden. Wie das Ergebnis aus dem Säulendiagramm aus Abbildung 6.10 zeigt, besteht bei den Teilnehmern beider Gruppen ein eher unterdurchschnittliches Wissen hinsichtlich der Produktgruppe „Full HD Beamer“.

Alle Teilnehmer der Rated Tags-Gruppe wurden ferner gefragt, wie sie die Bedienbarkeit und die Nützlichkeit von Rated Tags auf einer 7-Punkte Likert Skala bewerten würden. Die Ergebnisse zu diesen beiden Fragen sind Abbildung 6.11 zu entnehmen. Wie die Abbildung zeigt, bewerteten die meisten Teilnehmer die Bedienbarkeit mit einfach bis sehr einfach. Die Nützlichkeit von Rated Tags wurden von den Rated Tags-Teilnehmern häufig als sehr hilfreich bewertet.

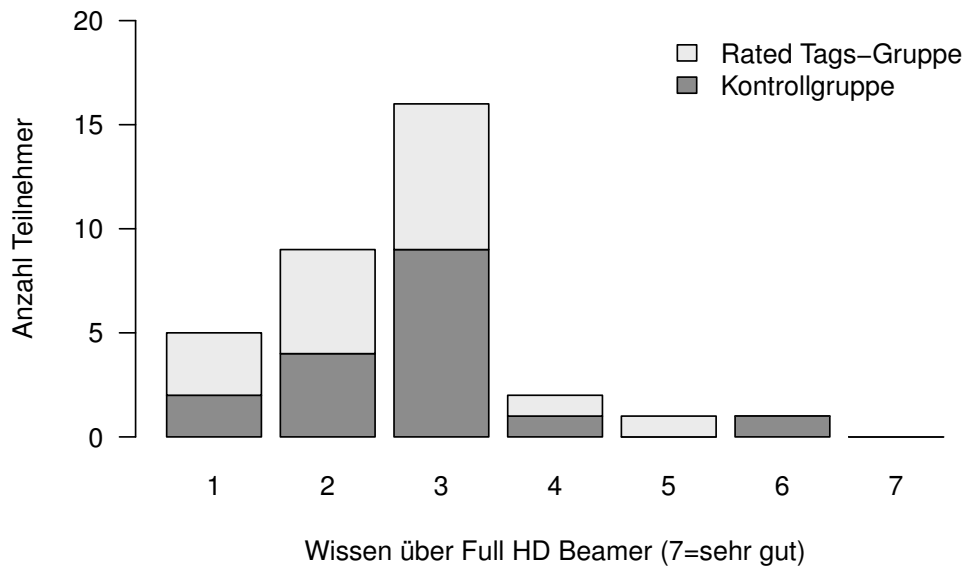
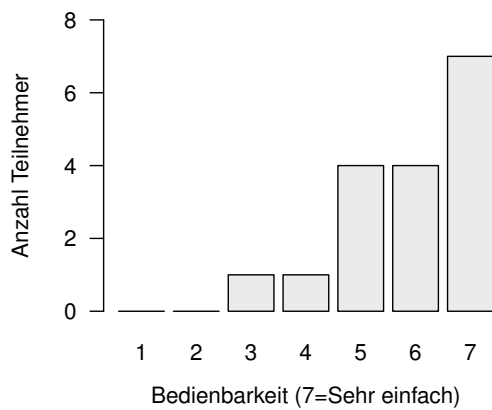
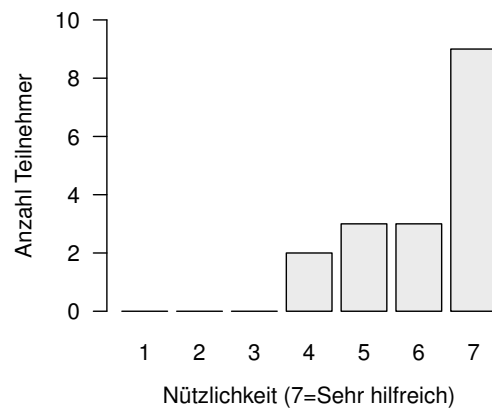


Abbildung 6.10: Produktgruppenwissen der Teilnehmer hinsichtlich der Produktgruppe „Full HD Beamer“ basierend auf einer 7-Punkte Likert Skala (N=34)



(a) Bedienbarkeit von Rated Tags aus Sicht der Rated Tags-Teilnehmer auf einer 7-Punkte Likert Skala (N=17)



(b) Nützlichkeit von Rated Tags aus Sicht der Rated Tags-Teilnehmer auf einer 7-Punkte Likert Skala (N=17)

Abbildung 6.11: Antworten der Teilnehmer bezüglich Bedienbarkeit und Nützlichkeit von Rated Tags

Abschließend wurden die Rated Tags-Teilnehmer noch befragt, ob es hinsichtlich des Ansatzes Probleme gab oder ob sie Verbesserungsvorschläge hätten. Eine Antwort im Freitext wurde jedoch nur von 3 der 17 Rated Tags-Teilnehmer abgegeben. Ein Teilnehmer hat einen interessanten Vorschlag zur Erweiterung des Rated Tags-Ansatzes genannt. Und zwar wurde vorgeschlagen, das Vorkommen der Rated Tags innerhalb der Rezension in irgendeiner Weise hervorzuheben, um die entsprechende Textstelle an der ein Rated Tag diskutiert wird schneller zu finden. Alle Anmerkungen der Teilnehmer sind in Tabelle B.2 auf Seite 224 zu finden.

6.3 Diskussion

Nach der Präsentation der Ergebnisse im vorigen Abschnitt sollen in diesem Abschnitt deren Diskussion und Interpretation stattfinden. Ferner werden die Limitationen die mit der Studie einhergehen diskutiert und bewertet. Abschließend werden mögliche Implikationen der Studie sowohl für die Forschung als auch für den praktischen Einsatz erläutert.

6.3.1 Interpretation der Ergebnisse

Wie die Ergebnisse aus dem vorigen Abschnitt gezeigt haben, konnte die Web-basierte Entscheidungshilfe Rated Tags die Entscheidung von Teilnehmern positiv beeinflussen. Teilnehmer der Rated Tags-Gruppe hatten eine signifikant erhöhte Wahrscheinlichkeit das dominierende Produkt auszuwählen. Ferner wurde der zeitliche Entscheidungsaufwand, also die benötigte Zeit um eine Entscheidung zu fällen, für Rated Tags-Teilnehmer signifikant reduziert. Die Studienergebnisse unterstützen folglich also die aufgestellten Hypothesen H1 und H2. Die Steigerung der Entscheidungsqualität sowie die Reduzierung des Entscheidungsaufwands entspricht nicht der traditionellen Auffassung von Payne et al., dass Entscheidungsträger einen Kompromiss zwischen diesen beiden Zielen eingehen müssen [154]. Dies zeigt, dass die Entscheidungshilfe Rated Tags das Potenzial zur Verbesserung der Produktsuche und Kaufentscheidung von Konsumenten besitzt.

Im Hinblick auf die Sicherheit das korrekte Produkt gewählt zu haben, waren Rated Tags-Teilnehmer nur geringfügig sicherer in ihrer Entscheidung. Die aufgestellte Hypothese H3 wird daher durch die Studienergebnisse nicht unterstützt und muss verworfen werden. Eine mögliche Erklärung für die geringen Unterschiede hinsichtlich der Entscheidungssicherheit könnte sein, dass Personen generell dazu tendieren in ihrer Entscheidung zu selbstsicher zu sein [94], was die Wirkung von Rated Tags abgeschwächt haben könnte. Eine Möglichkeit könnte auch das fehlende Vertrauen in die interaktive Entscheidungshilfe Rated Tags sein. Eine Studie von Vijayasathy und Jones hat beispielsweise einen negativen Zusammenhang zwischen Entscheidungshilfen und der Entscheidungssicherheit von Personen festgestellt und als Grund dafür das fehlende Vertrauen in die Entscheidungshilfe identifiziert [201]. Eine Übertragung dieser Erkenntnis auf die Rated Tags-Studie könnte bedeuten, dass Teilnehmer möglicherweise kein bzw. ein geringes Vertrauen in die aggregierte Bewertung der Rated Tags hatten. Aus den qualitativen Ergebnissen der Studie konnte ein solches fehlendes Vertrauen allerdings nicht abgeleitet werden.

Ein sehr interessantes Ergebnis aus Abschnitt 6.2.3 ist, dass die Gesamtbewertungen der Rezensionen im Mittel signifikant von den durchschnittlichen Bewertungen der Rated Tags abweichen. Dieses Ergebnis verdeutlicht den Bedarf einer Entscheidungshilfe wie Rated Tags, da ersichtlich ist, dass die Gesamtbewertung der Rezensenten durch eine Gewichtung der Produktaspekte zustande kommt. Da jedoch verschiedene Konsumenten verschiedene Anforderungen haben und damit Produktaspekte unterschiedlich gewichten, ist eine Gesamtbewertung für die Entscheidungsfindung wenig hilfreich. Mit Hilfe des Rated Tags-Ansatzes sehen Konsumenten die einzelnen, ungewichteten Produktaspekte sowie deren Bewertung und können diese dann entsprechend ihren eigenen Anforderungen gewichten.

Wie der Fragebogen gezeigt hat, waren fast alle Rated Tags-Teilnehmer der Meinung, dass Rated Tags alleine, d.h. ohne Rezensionen, nicht für eine finale Entscheidung ausreichen würden. Dies deckt sich ebenfalls mit den Ergebnissen von Chevalier und Mayzlin, welche herausgefunden haben, dass sich Konsumenten nicht allein auf aggregierte Bewertungen verlassen, sondern zusätzlich Kundenrezensionen heranziehen [29]. Der Rated Tags-Ansatz ist daher nicht als Ersatz, sondern als Ergänzung zu traditionellen Rezensionen gedacht.

6.3.2 Limitationen der Studie

Die allgemeine Übertragbarkeit der Ergebnisse unterliegt einigen Einschränkungen, welche nachfolgend näher adressiert werden sollen. Eine Einschränkung betrifft die Studienteilnehmer. Die Fragebogenergebnisse haben gezeigt, dass die Teilnehmer zur potenziellen Anwendergruppe von Rated Tags gehören, da alle Teilnehmer bereits Rezensionen verwendet haben und ihnen auch eine sehr hohe Bedeutung zuordnen. Jedoch war der Umfang der verwendeten Stichprobe mit 34 Teilnehmern relativ gering. Des Weiteren zeigt die Altersverteilung der Teilnehmer, dass primär jüngere Personen von 18 bis 35 Jahre an der Studie beteiligt waren. Eine Anwendbarkeit der Ergebnisse auf Personen, die dieses Alter übersteigen, ist basierend auf der Studie nicht möglich.

Ferner lag der primäre Fokus der Studie in der Betrachtung der Entscheidungsunterstützung und weniger in der Betrachtung der erzeugten Rated Tags. Entgegen der späteren Nutzung von Rated Tags, bei der Anwender ihre eigenen Rezensionen mit Rated Tags versehen, erfolgte in der Studie die Zuweisung von Rated Tags an Rezensionen durch ausgewählte Produktexperten. Die Studie kann daher keine Aussage über die Erzeugung von Rated Tags durch echte Konsumenten treffen.

Eine weitere Einschränkung der Studie ist, dass nur eine bestimmte Produktgruppe („Full HD Beamer“) herangezogen wurde. Gemäß Nelson lassen sich Produkte in Such- und Erfahrungsgüter einteilen [139]. Die Produktgruppe der Full HD Beamer enthält sowohl Such- als auch Erfahrungseigenschaften. Sucheigenschaften sind objektiv vergleichbare Merkmale die auch ohne Rezensionen ermittelt werden können [171, S. 141], wie etwa der Preis oder die maximale Auflösung in Pixel. Mögliche Erfahrungseigenschaften betreffen beispielsweise die Bildqualität oder die subjektive Einschätzung des Betriebsgeräusches. Da andere Bewertungsobjekte auch ein anderes Gleichgewicht an Such- und Erfahrungseigenschaften aufweisen, ist eine Pauschalisierung hier nicht möglich. Die Be-

wertung eines Hotels hat beispielsweise einen größeren Anteil an Erfahrungseigenschaften, während die Anzahl der objektiven Sucheigenschaften (z.B. die Anzahl der Hotelsterne) eher gering ist.

Ferner wurden in der Studie bewusst nicht alle Phasen des Entscheidungsfindungsprozesses betrachtet. Die Annahmen der Studie waren, dass ein Kunde bereits einen relevanten Onlineshop ausgewählt hat und dass die Menge der infrage kommenden Alternativen auf fünf begrenzt wurde. Daher kann die Studie keine Aussage über das Verhalten zur Selektion der Alternativen liefern. Die geringe Anzahl an fünf Produktalternativen kann ebenfalls als Einschränkung der Studie angesehen werden, da hier noch eine recht hohe Wahrscheinlichkeit besteht, dass ein Teilnehmer das dominierende Produkt zufällig und nicht aufgrund einer sorgfältigen Analyse gewählt hat.

In der Studie wurden je Produkt nur 20 Rezensionen herangezogen. In beliebten Onlineshops ist die Anzahl der Rezensionen je Produkt jedoch meist deutlich höher. Eine höhere Anzahl an Rezensionen könnte dazu führen, dass auch mehr verschiedenartige und womöglich unbrauchbare Tags erzeugt werden. Obwohl eine gestiegene Anzahl an Tags aufgrund der gestiegenen Informationen auch eine positive Auswirkung auf die Entscheidungsfindung haben kann, besteht jedoch auch das Risiko, dass die Komplexität der Entscheidung ansteigt.

Abschließend muss erwähnt werden, dass die Anweisung zur Selektion des dominierenden Produkts bei den Teilnehmern eine implizite Einschränkung hinsichtlich der verwendeten Entscheidungsstrategie mit sich gebracht hat. Da es die Aufgabe war ein Produkt anhand nur eines dominierenden Merkmals („Betriebsgeräusch“) auszuwählen, wurde den Teilnehmern indirekt die lexikographische Entscheidungsheuristik nahegelegt. Die Wahl nur eines dominierenden Merkmals hat es den Teilnehmern einfacher gemacht sich zu entscheiden [41, S. 273]. Allerdings werden für reale Kaufentscheidungen oft viel komplexere Entscheidungsstrategien verwendet, welche jedoch im Rahmen der Studie nicht abgebildet werden konnten. Hinzu kommt, dass die Teilnehmer der Studie ein relativ geringes Produktgruppenwissen hatten (siehe Abbildung 6.10). Es besteht die Möglichkeit, dass Teilnehmer mit einem höheren Produktgruppenwissen ein anderes Entscheidungsverhalten aufgewiesen hätten als die untersuchte Stichprobe.

6.3.3 Implikationen für die Forschung

Basierend auf den Ergebnissen der Studie ergeben sich einige Implikationen für die Forschung. Dies betrifft beispielsweise den Forschungsbereich des Social Tagging. Wie in der Studie gezeigt werden konnte, waren die Produktexperten in der Lage sinnvolle bewertbare Tags auf einer vorgegebenen Bewertungsskala zu erstellen. Vig et al. haben in ihrer Arbeit vermutet, dass die Kombination eines Tags mit einer solchen mehrwertigen Bewertungsskala zu komplex für Anwender sei und einen zu hohen kognitiven Aufwand darstellen würde [200, S. 325]. Basierend auf den Studienergebnissen trifft diese Vermutung weder für die Erstellung noch für die Verwendung der Tags zu. Dies ist eine wichtige Erkenntnis für verwandte Forschungsarbeiten, die sich mit der Kombination von Tags und Bewertungen befassen.

Aus dem qualitativen Feedback der Teilnehmer hat sich eine interessante Erweiterung für zukünftige Forschungsarbeiten ergeben. Und zwar wurde von einem Teilnehmer vorgeschlagen, die Position des vorkommenden Rated Tags im Text zu markieren, um (vor allem in einem längeren Text) den diskutierten Aspekt schneller zu finden. Basierend auf diesem Vorschlag wurde ein Konzept entworfen, prototypisch umgesetzt und publiziert [88]. Eine Empfehlung für weitere Forschungsarbeiten ist die Evaluierung dieses Ansatzes. Weitere Informationen hierzu finden sich im Ausblick dieser Arbeit (siehe Kapitel 9).

Die Ergebnisse der Studie könnten auch Einflüsse auf den Forschungsbereich der Empfehlungssysteme ausüben. Rated Tags bietet dank der Benutzer-generierten Tags eine Möglichkeit zur Erstellung von detaillierteren Bewertungen. Wie Adomavicius und Kwon beschreiben, wird die Bedeutung von detaillierten Bewertungen (zusätzlich zu einer Gesamtbewertung) im Bereich der Empfehlungssysteme zukünftig ansteigen [1]. Gedikli und Jannach erwähnen hierzu weiterhin, dass detaillierte Bewertungen eine wertvolle Quelle sind, um den Konsumenten besser zu verdeutlichen, wieso ihnen bestimmte Empfehlungen präsentiert wurden [51]. Rated Tags könnten somit zusätzlich zu den Gesamtbewertungen eine wichtige Basis für die Erforschung von neueren Empfehlungssystemen bilden.

Die Ergebnisse der Studie sind vielversprechend, so dass eine Anpassung der Studie für zukünftige Forschungsarbeiten naheliegt. Hierzu wäre insbesondere die Abdeckung früherer Phasen der Entscheidungsfindung ratsam, um die Nützlichkeit von Rated Tags in diesen Bereichen zu evaluieren. Ferner sollten hier die verschiedenen Einschränkungen der Studie adressiert werden, z.B. die Untersuchung einer Teilnehmergruppe mit einem höheren Produktgruppenwissen. Wie Kamis und Davern ermittelt haben, führt ein erhöhtes Produktgruppenwissen der Anwender zu einer erhöhten empfundenen Bedienbarkeit und Nützlichkeit einer Entscheidungshilfe [94]. Diese Korrelation sollte in Bezug auf Rated Tags weiter evaluiert werden. Ferner wäre eine genauere Akzeptanzüberprüfung von Rated Tags aus technologischer Sicht sinnvoll, z.B. mithilfe des *Technology Acceptance Models (TAMs)* von Davis [39].

6.3.4 Implikationen für die Praxis

Obwohl sich die Auswirkungen des Rated Tags-Ansatzes primär auf die Konsumenten beziehen, können jedoch auch Onlinehändler von einer Implementierung des Ansatzes profitieren. Anhand einer Analyse der durch die Benutzer erzeugten Detailbewertungen lassen sich Benutzerprofile noch genauer erstellen als über die Analyse von Gesamtbewertungen. Ferner könnten sich die durch Rated Tags verminderte Entscheidungsdauer und die einfachere Auffindbarkeit von relevanten Produktinformationen positiv auf die Kundenzufriedenheit auswirken, was letztendlich auch dem Onlinehändler zugutekommt.

Die Studie hat gezeigt, dass Anwender in der Lage waren, sinnvolle Produktmerkmale in Form von Rated Tags zu erstellen. Basierend auf den generierten Rated Tags, eröffnen sich für Onlinehändler zusätzliche Analysemöglichkeiten auf Produkt- und Produktgruppenebene. So lassen sich beispielsweise auf Produktgruppenebene diejenigen Merkmale identifizieren, die von den Anwendern am häufigsten verwendet wurden und somit für eine Produktgruppe die größte Relevanz aufweisen. Im Hinblick auf die Produktebene

könnten Onlinehändler die generierten Merkmale analysieren, um einfach herauszufinden, welche Stärken oder Schwächen ein Produkt aus Kundensicht aufweist.

Im Hinblick auf die Qualität der erzeugten Rated Tags wurde ermittelt, dass 7 der 57 einzigartigen Tags eine geringe Qualität aufweisen und somit auch die Qualität der Entscheidungshilfe mindern. Die drei Problembereiche qualitativ geringwertiger Tags waren Rechtschreibfehler, geringe Relevanz sowie semantische Redundanz einiger Tags. Für die Adressierung von Rechtschreibfehlern sollte eine Rechtschreibkomponente integriert werden, welche die Tags der Anwender noch vor dem Speichern analysiert und mögliche Rechtschreibkorrekturen vorschlägt. Die verminderte Qualität durch nicht relevante Tags ist schwer adressierbar. Eine Möglichkeit wäre die Anpassung der Dokumentation, um Anwendern zu verdeutlichen, bei welchen Tag-Arten es sich um hilfreiche Tags handelt. Um die Anzahl an semantisch ähnlichen Tags zu reduzieren sind verschiedene Ansätze denkbar. Ein Ansatz wäre ein kollaboratives Votingsystem, das die Zusammenführung von ähnlichen Tags durch die Anwender erlaubt. Eine beispielhafte Umsetzung hierzu liefert die Webseite LibraryThing¹. Ein weiterer Ansatz zur teilautomatisierten Vereinheitlichung der Tags wird im nächsten Kapitel vorgestellt.

¹http://librarything.com/wiki/index.php/Tag_combining (besucht am 25.06.2014)

Kapitel 7

Vereinheitlichung der erstellten Tags

Im vorigen Kapitel wurde der Rated Tags-Ansatz im Rahmen einer Anwenderstudie evaluiert. Um den Mehrwert von Rated Tags hinsichtlich der Entscheidungsunterstützung von Konsumenten weiter zu steigern, soll in diesem Kapitel ein Modell für eine teilautomatisierte Vereinheitlichung der erzeugten Tags erstellt und evaluiert werden. Hierzu werden zunächst die angrenzenden Themengebiete und Ansätze zur teilautomatisierten Gruppierung von Produktaspekten näher betrachtet. Danach wird ein Klassifikationsansatz vorgestellt, der die speziellen Charakteristiken des Rated Tags-Ansatzes berücksichtigt. Die Leistung des Klassifikators wird dann im Rahmen einer Untersuchung evaluiert. Die nachfolgenden Ausführungen basieren dabei zu einigen Teilen auf dem Artikel „Grouping Product Aspects from Short Texts Using Multiple Classifiers“ [91].

7.1 Problemanalyse und Zielsetzung

Wie bereits in den vorangegangenen Kapiteln erläutert wurde, basiert der Rated Tags-Ansatz auf den Konzepten des kollaborativen Taggings. Dies bedeutet, dass jeder Anwender in der Lage ist, Benutzer-definierte Tags zu erstellen. Da die Inhaltsklassifikation nicht durch eine zentrale Kontrollinstanz verwaltet wird, besteht die Gefahr, dass die Anwender verschiedenartige Tags erstellen, welche sich jedoch auf den gleichen Aspekt beziehen. Um dem entgegen zu wirken, bietet das Rated Tags-Konzept die Wiederverwendbarkeit von bestehenden Tags sowie die Vordefinition von wichtigen Tags durch den Betreiber an. Allerdings kann auch dadurch nicht komplett verhindert werden, dass von Anwendern semantisch ähnliche Tags durch verschiedenartige Schlagwörter ausgedrückt werden. Ein Beispiel wären die Tags „toller Klang“ und „schlechter Sound“, welche beide den Aspekt „Klangqualität“ umschreiben.

Die oben genannte Problematik führt zu einer Erhöhung der Entscheidungskomplexität für Konsumenten, da ein erhöhter kognitiver Aufwand notwendig ist, um die semantisch gleichen Informationen anhand der unterschiedlichen Tags zu vergleichen. Dies soll an dem Beispiel in Abbildung 7.1 demonstriert werden. Die Abbildung zeigt verschiedene Benutzer-generierte Pro- und Kontra-Punkte zu einem bestimmten Produkt¹. Da sich

¹Die Daten stammen aus dem Bewertungsportal Buzzillions, welches laut eigenen Angaben mehr als 17 Millionen Rezensionen umfasst (siehe <http://www.buzzillions.com/about>, besucht am 31.01.2015)

Great picture quality (56)	Sound (17)
Picture (33)	Speakers (5)
Everything (13)	Poor audio (4)
3d (9)	Sound quality (4)
Price (9)	The sound (3)
Features (6)	Price (3)

Abbildung 7.1: Benutzer-generierte Erzeugung von ähnlichen Aspekten durch verschiedenartige Schlagwörter (Quelle: <http://www.buzzillions.com>, besucht am 31.01.2015)

verschiedene Punkte auf den gleichen Aspekt beziehen (z.B. „Sound“ und „Poor audio“), ist es für den Endanwender schwer, einen einheitlichen Überblick über die Produktaspekte zu erlangen.

Die aus Abbildung 7.1 ersichtliche Problematik trifft ebenfalls auf den Rated Tags-Ansatz zu. Das weitere Ziel dieser Arbeit ist es daher ein Modell zu entwickeln, das die Benutzer-generierten Tags in einer teilautomatischen Weise vereinheitlicht. Da es sich bei den Tags letztendlich um Produktmerkmale handelt, kann das Problem als Identifikation und Gruppierung von Produktmerkmalen bzw. Produktaspekten klassifiziert werden. Neben der Gruppierung ist auch eine Erkennung von irrelevanten Inhalten notwendig, um nicht aussagekräftige Inhalte bzw. Inhalte, die kein konkretes Produktmerkmal adressieren, zu filtern.

Die Zielsetzung für die Überführung von Benutzer-generierten Tags in einheitliche Produktaspekte ist in Abbildung 7.2 dargestellt. Wie die Abbildung zeigt, werden wichtige Produktaspekte einer Produktgruppe durch den Betreiber vordefiniert. Das vom Betreiber erstellte und kontrollierte Vokabular kann als leichtgewichtige *Ontologie* aufgefasst werden. Gruber definiert eine Ontologie als eine „explizite Beschreibung einer Konzeptualisierung“ [54, S. 199]. Das von den Anwendern erzeugte nicht-kontrollierte Vokabular, also die erstellten Tags im Rahmen des Rated Tags-Modells, entspricht hingegen dem Konzept einer Folksonomie. Das weitere Ziel dieser Arbeit ist es, diese Folksonomie in die leichtgewichtige Ontologie des Betreibers zu überführen. Relevante Schlagwörter zu dieser Überführung sind „Ontology Matching“ oder „Ontology Learning“. Der Prozess zur Vereinheitlichung der Tags wird in Abbildung 7.2 auch als Informationsintegration bezeichnet.

Im weiteren Verlauf dieses Kapitels werden in Abschnitt 7.2 zunächst die für die Problemstellung relevanten Themengebiete kurz erläutert sowie der aktuelle Forschungsstand identifiziert und zur Zielsetzung dieses Kapitels abgegrenzt. In Abschnitt 7.3 wird dann ein Klassifikationsmodell entworfen, welches für das Ontology Matching eingesetzt werden soll. Dabei wird unter anderem ein Klassifikator aus dem Bereich des überwachten Lernens (engl. *supervised learning*) entwickelt. Eine Erprobung dieser Komponente wird dann abschließend in Abschnitt 7.4 im Rahmen einer Untersuchung durchgeführt. Dabei werden fünf verschiedene Produktgruppen betrachtet, um die Anwendbarkeit des neuen Klassifikators auf heterogene Produktgruppen zu demonstrieren.

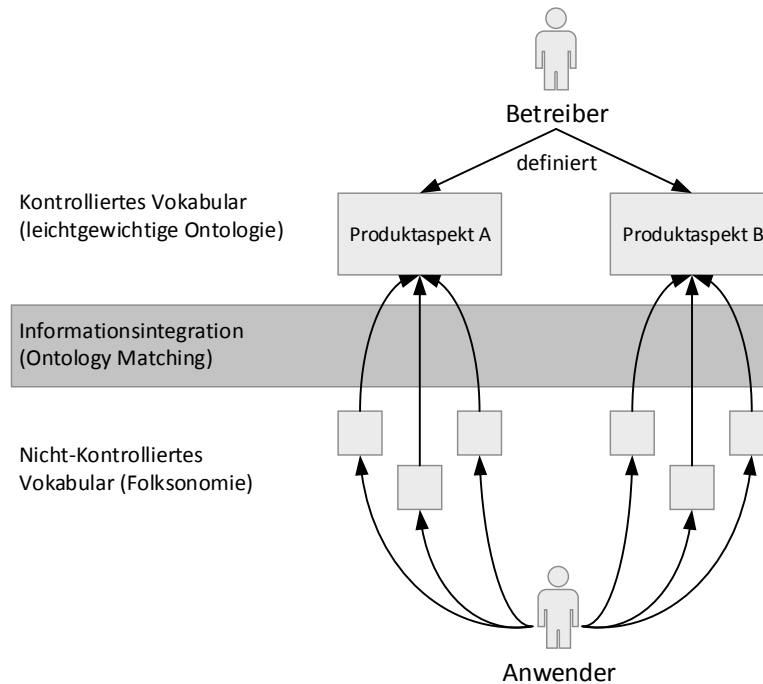


Abbildung 7.2: Überführung von Benutzer-generierten Tags in einheitliche Produktraspekte

7.2 Verwandte Arbeiten

Für die oben genannte Problemstellung sind diverse Themengebiete relevant, welche im Rahmen der nachfolgenden Unterabschnitte näher betrachtet werden sollen. Zunächst werden in Abschnitt 7.2.1 die angrenzenden Themenkomplexe aufgeführt sowie deren Bezug zur Problemstellung verdeutlicht. Anschließend werden die existierenden Forschungsansätze zur Lösung des Problems analysiert und zum Lösungsansatz dieser Arbeit abgegrenzt.

7.2.1 Einordnung

Da es sich bei der (teil-)automatisierten Vereinheitlichung der Tags um die Verarbeitung und Analyse von kurzen Texten (Tags) handelt, lässt sich als übergeordnetes Themengebiet der Bereich *Text Mining* identifizieren. Dieser Forschungsbereich umfasst gemäß Sebastiani alle Tätigkeiten der Textanalyse zur Gewinnung von relevanten Informationen [173, S. 2]. Im Speziellen ist für diese Arbeit der Bereich der *Textklassifikation* interessant, welcher laut Sebastiani als eine Ausprägung des Text Mining angesehen werden kann und unter anderem Berührungspunkte zu den Disziplinen des maschinellen Lernens (engl. *machine learning*) sowie der Informationsrückgewinnung (engl. *information retrieval*) aufweist [173, S. 2]. Bei der Textklassifikation (auch Textkategorisierung genannt) besteht die Aufgabe darin einen Klassifikationsmechanismus (kurz Klassifikator) zu entwerfen,

welcher zu jedem Textdokument die dazugehörige Klasse bzw. Kategorie ermittelt. Eine formale Darstellung der Textklassifikation wird im Rahmen des Klassifikationsmodells in Abschnitt 7.3.1 vorgenommen.

Wie bereits Sebastiani festgestellt hat, wird der Begriff der Textklassifikation in der Literatur teilweise unterschiedlich aufgefasst. Im Gegensatz zu der oben genannten Zuweisung eines Dokuments zu einer Klasse wird der Begriff auch für die automatische Identifikation der Klassen sowie deren Gruppierung verwendet, was jedoch üblicherweise als *Text Clustering* bezeichnet wird [173, S. 2]. Eine solche automatisierte Gruppierung von Aspekten im Sinne eines Text Clustering kommt im Rahmen dieser Arbeit jedoch nicht infrage, da der Abstraktionsgrad hinsichtlich der Kategorisierung von Aspekten von den jeweiligen Bereitstellern des Rated Tags-Systems definierbar sein muss. Dies ist darin begründet, dass verschiedene Anbieter eine unterschiedliche Kategorisierung vornehmen. So wird beispielsweise ein Onlineshop, der sich auf Digitalkameras spezialisiert hat, eine wesentlich feinere Kategorisierung hinsichtlich der Kameramerkmale vornehmen, als ein Onlineshop, dessen Sortiment neben Kameras auch weitere verschiedenartige Elektroprodukte umfasst. Die Notwendigkeit einer Benutzerbeteiligung bei der Gruppierung von Synonymen wird ferner durch Zhai et al. hervorgehoben, welche ebenfalls die Gruppierung von Produktmerkmalen basierend auf Rezensionen untersucht haben [219, S. 1272]. Auch Carenini et al. postulieren, dass bei der Kategorisierung von Benutzermeinungen in Freitextform die Integration von bestehendem Wissen über das zu bewertende Objekt notwendig ist [25, S. 18].

Da es sich bei den zu klassifizierenden Dokumenten, also den einzelnen Rated Tags, um Beschreibungen von Meinungen zu Produktaspekten handelt, ist ebenfalls der Forschungsbereich des Opinion Mining bzw. des Aspekt-orientierten Opinion Mining von Interesse. Popescu und Etzioni unterteilen das Problem des Opinion Minings in die folgenden Unteraufgaben [159, S. 9]:

1. Identifikation von Produktaspekten
2. Identifikation von Meinungen zu Produktaspekten
3. Bestimmung der Polarität der Meinungen, d.h., handelt es sich um eine positive oder negative Meinung?
4. Bestimmung der Ausprägung der Polarität

Im Rahmen des Rated Tags-Ansatzes sind nur die ersten beiden Punkte relevant, d.h. die Identifikation und Extraktion von Produktaspekten und deren Meinungen. Die Punkte 3 und 4 stellen im Gegensatz zum Opinion Mining aus Rezensionen keine Probleme dar. Dies liegt daran, dass die Polarität (Punkt 3) sowie die Stärke der Meinungen (Punkt 4) explizit durch die mehrwertige Bewertungsskala eines Rated Tags zum Ausdruck gebracht werden.

Der Bereich des Aspekt-orientierten Opinion Mining befasst sich insbesondere mit der Analyse von Benutzermeinungen im Hinblick auf spezielle Aspekte. Liu et al. haben diesbezüglich die drei am Häufigsten im Web vertretenen Rezensionsformate zur Bewertung von Aspekten identifiziert [119, S. 343]:

- Format F_{PK} - Pro und Kontra: Die Rezensenten haben die Möglichkeit Pro- und Kontrapunkte für einzelne Aspekte zu vergeben.
- Format F_D - Freitext-Rezension: Rezensenten können eine Rezension im Freitextformat verfassen, aber keine Pro- und Kontra-Punkte für Aspekte vergeben.
- Format F_{PKD} - Pro/Kontra und Freitext-Rezension: Zusätzlich zum Format F_{PK} sind die Rezensenten in der Lage eine detaillierte Rezension im Freitextformat zu verfassen.

Der Rated Tags-Ansatz lässt sich am ehesten mit den Formaten F_{PK} und F_{PKD} vergleichen. Anstelle der Pro- und Kontrapunkte wird allerdings eine mehrwertige Bewertungsskala verwendet. Die Problematik des Rated Tags-Ansatzes hinsichtlich der Identifikation und Aggregation von ähnlichen Aspekten ist jedoch identisch zur Problematik der Formate F_{PK} und F_{PKD} . Daher werden in Abschnitt 7.2.2 einige Arbeiten aus diesem Bereich betrachtet. Liu et al. postulieren, dass die Analyse der Pro- und Kontrapunkte genauere Ergebnisse liefert als die Analyse der detaillierten Rezension [119, S. 343]. Aus diesem Grund beschränkt sich das Klassifikationsmodell sowie die spätere Evaluierung auf das Rezensionsformat F_{PK} .

Für die Identifikation von Aspekten ist ferner zu beachten, dass die Beschreibung von Aspekten sowohl explizit als auch implizit geschehen kann [42, 119, 159, 189]. Bei einer expliziten Beschreibung eines Aspekts ist der Name des Aspekts in der Beschreibung enthalten. So enthält beispielsweise die Beschreibung „Teurer Preis“ explizit den Aspekt „Preis“. Hingegen würde die Beschreibung „Zu teuer“ den Aspekt „Preis“ nicht enthalten, sondern nur implizieren. Für Su et al. ist ein impliziter Aspekt dadurch gekennzeichnet, dass er 1) den Aspekt nicht explizit enthält und 2) durch eine Betrachtung der ihn umgebenden Wörter der Rezension in einen expliziten Aspekt überführt werden kann [189, S. 961]. Letzteres Merkmal ist jedoch im Rahmen des Rated Tags-Ansatzes nicht anwendbar, da hier Aspekte häufig nur mit einem Wort oder wenigen Wörtern beschrieben werden, wodurch eine Kontextermittlung, und damit eine Rückführung von einem impliziten zu einem expliziten Aspekt, deutlich erschwert wird.

Bei dem Prozess der Aspektgruppierung sind aus konzeptioneller Sicht unter anderem Ansätze der Textklassifikation sowie des Text Clusterings relevant. Eine nähere Betrachtung dieser Themenfelder wird im Rahmen des Klassifikationsmodells (siehe Abschnitt 7.3.1) vorgenommen. Eine zusammenfassende Darstellung zur Einordnung des Themengebietes ist in Abbildung 7.3 zu finden, wobei sich der Schwerpunkt der Arbeit mit der Gruppierung von Aspekten beschäftigen wird.

7.2.2 Abgrenzung

Nach der Einordnung in die angrenzenden Themengebiete soll in diesem Abschnitt nun eine Abgrenzung zu den existierenden Arbeiten aus den Bereichen der Aspektidentifikation und Aspektgruppierung vorgenommen werden. Letzteres ist besonders wichtig, um die Anwender bei ihrer Entscheidung angemessen zu unterstützen. Ohne eine Gruppierung der Aspekte werden meist zu viele, teilweise überlappende Aspekte identifiziert. Ein Beispiel

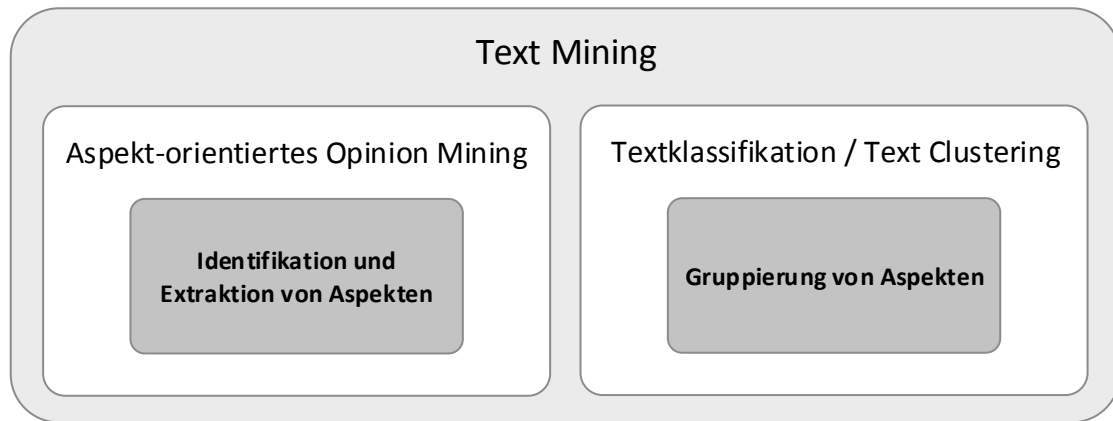


Abbildung 7.3: Grobe Einordnung des Klassifikationsproblems in verwandte Themengebiete

ist der in [73] verwendete Datensatz¹, bei dem vier der fünf untersuchten Produkte mit über 100 unterschiedlichen Aspekten annotiert wurden, was bei Anwendern zu einer gestiegenen Entscheidungskomplexität führt. In ihrem Survey-Artikel zu Opinion Mining nennen Liu und Zhang jedoch lediglich die zwei Ansätze von Carenini et al. [25] und Zhai et al. [219] zur Gruppierung von Aspekten [120, S. 31 f.]. Diese und weitere Ansätze sollen in diesem Abschnitt zum Ansatz dieser Arbeit abgegrenzt werden.

Eine der frühen Arbeiten zur Extraktion von Aspekten aus Rezensionen (Format F_D) sowie der damit verbundenen Sentimentanalyse der enthaltenen Sätze stammt von Hu und Liu [73, 74]. Zum Einsatz kommt ein Part of speech (POS)-Tagging des Rezensionstextes zur Ermittlung der Wortarten. Die Argumentation von Hu und Liu ist, dass Produktaspekte primär als Substantive ausgedrückt werden, d.h., die erzeugten POS-Tags werden nach Substantiven gefiltert, um Produktaspekte zu ermitteln. Eine weitere These ihrer Forschungsarbeit ist, dass häufig vorkommende Substantive mit höherer Wahrscheinlichkeit Produktaspekten entsprechen als Substantive, die seltener vorkommen. Es wird daher eine Assoziationsanalyse mithilfe des Apriori-Algorithmus durchgeführt, um die häufiger vorkommenden Substantive zu ermitteln. Da der Ansatz nur Substantive betrachtet, wird hier keine Berücksichtigung von impliziten Aspekten vorgenommen [73, S. 171].

Eine Wiederverwendung des Ansatzes von Hu und Liu [74] wird durch Carenini et al. vorgenommen [25]. Sie erweitern den Ansatz um den Einsatz einer Produktgruppentaxonomie, um bestehendes Produktgruppenwissen in den Algorithmus einfließen zu lassen. Allerdings ist aus deren beispielhaften Ausschnitt der Taxonomie für Digitalkameras ersichtlich, dass auch Fakten-basierte Merkmale (z.B. Anzahl der effektiven Pixel) betrachtet werden, zu welchen typischerweise keine Meinungen abgegeben wird. Die zweite Schlüsselkomponente des Ansatzes verwendet verschiedene Ähnlichkeitsmetriken. Es wird

¹<http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip> (besucht am 20.02.2015)

versucht die aus dem Ansatz von Hu und Liu ermittelten Aspekte auf die Elemente der Produktgruppentaxonomie zu überführen. Eine Evaluierung findet für die Produktgruppen „Digitalkameras“ und „DVD-Player“ statt. Für die Gruppierung der Aspekte wird der nicht-überwachte Ansatz von Hu und Liu [73] sowie ein überwachter Ansatz (basierend auf Produktgruppentaxonomien) miteinander kombiniert. Zudem wird für das Mapping auf die lexikalische Datenbank WordNet zugegriffen, um die Ähnlichkeiten zweier Begriffe zueinander zu ermitteln. Ein Kritikpunkt des Ansatzes ist, dass nicht für jede Produktgruppe eine Taxonomie existiert, wodurch die Wiederverwendbarkeit des Ansatzes eingeschränkt und ein manueller Aufwand zur Erstellung einer solchen Taxonomie erforderlich wird.

Der Ansatz von Liu et al. betrachtet die Identifikation von Aspekten aus Benutzer-generierten Pro- und Kontrapunkten (Format F_{PK}) [119]. Die dort vorgeschlagene Methode basiert auf *Natural Language Processing (NLP)* und verwendet supervised Rule Mining. Die Ergebnisse von Liu et al. zeigen, dass der oben genannte Ansatz von Hu und Liu ([73]) nur für komplette Sätze, d.h. für die Formate F_D und F_{PKD} , anwendbar ist, jedoch nicht für das Format F_{PK} , da Anwender hier keine ganzen Sätze formulieren [119, S. 343]. Eine Anwendung des Ansatzes von Liu et al. auf die Gruppierung von Synonymen erzielt mit einem F_1 -Wert von 68,42 % ein eher mittelmäßiges Ergebnis. Ein Kritikpunkt dieses Ansatzes ist, dass aus dem Artikel nicht hervorgeht wie viele Rezensionen manuell kategorisiert wurden („by manually labeling a large number of reviews“ [119, S. 346]) und somit der manuelle Aufwand nicht eingeschätzt werden kann, um die im Artikel präsentierten Ergebnisse zu erzielen. Ferner kritisieren Hu und Liu die Komplexität der von Liu et al. vorgestellten Methode [75, S. 62].

Hu und Liu untersuchen ebenfalls die Identifikation von Aspekten aus Benutzer-generierten Pro- und Kontrapunkten (Format F_{PK}) [75]. Sie verwenden supervised Class Sequential Rule Mining zur Extraktion von Aspekten anhand von lexikalischen Mustern. Die Ergebnisse ihres Ansatzes sind vergleichbar mit den Ergebnissen von Liu et al. [119], jedoch heben Hu und Liu im Vergleich dazu den geringeren manuellen Aufwand ihres Ansatzes hervor [75, S. 66]. Eine Betrachtung der Gruppierung von Aspekten findet in diesem Ansatz nicht statt.

Die Arbeit von Popescu und Etzioni untersucht Rezensionen (Formate F_D) und betrachtet dabei die Extraktion von expliziten und impliziten Aspekten [159]. Im Vergleich zur Arbeit von Hu und Liu erzielt der Ansatz eine deutlich höhere Präzision bei der Aspektextraktion [159, S. 10]. Ausschlaggebend hierfür ist die Verwendung von Web-basiertem Pointwise Mutual Information (PMI) [159, S. 13]. PMI wird grundsätzlich dazu verwendet, um den Zusammenhang zwischen zwei Zufallsvariablen zu ermitteln und kann gemäß Turney wie folgt dargestellt werden [198, S. 2]:

$$PMI(x, y) = \log_2 \left(\frac{p(x, y)}{p(x) \cdot p(y)} \right) \quad (7.1)$$

In Gleichung 7.1 stehen $p(x)$ und $p(y)$ jeweils für die Wahrscheinlichkeit, dass ein Wort x bzw. y vorkommt. Die Wahrscheinlichkeit $p(x, y)$ steht für das gleichzeitige Vorkommen beider Wörter. Ist das Vorkommen beider Wörter komplett unabhängig, so

gilt $p(x, y) = p(x) \cdot p(y)$, wodurch sich $\log_2 1$, also der Wert 0 ergibt. Das in Popescu und Etzioni genannte Web PMI adaptiert PMI so, dass anstelle der Wahrscheinlichkeiten die Trefferanzahl eines Wortes basierend auf einer Suchmaschinenanfrage verwendet wird. Die nachfolgende Gleichung zeigt den Ansatz von Web PMI, wobei $Hits$ für eine Funktion steht, welche die Anzahl der Suchtreffer für die jeweiligen Wörter x und y zurückliefert [159, S. 12]:

$$Web\ PMI(x, y) = \frac{Hits(x, y)}{Hits(x) \cdot Hits(y)} \quad (7.2)$$

Obwohl der Ansatz von Popescu und Etzioni gegenüber dem Ansatz von Hu und Liu eine deutliche Steigerung der Präzision liefert, muss die Praxistauglichkeit des Ansatzes infrage gestellt werden. Zum einen besteht eine Abhängigkeit zu den Ergebnissen einer Suchmaschine und zum anderen sind bei diesem Ansatz hohe Laufzeitkosten zu erwarten, da für jede Ermittlung eines Web-basierten PMI-Wertes jeweils drei Anfragen an eine Suchmaschine gestellt werden müssen.

Ein weiterer Ansatz von Guo et al. verwendet das Format F_{PKD} , d.h. Pro- und Kontrapunkte sowie die dazugehörigen Rezensionen [56]. Ähnlich zu einigen vorangegangenen Ansätzen wird POS-Tagging verwendet um Substantive zu ermitteln, welche dann als Aspektkandidaten vermerkt werden. Im Gegensatz zu bereits genannten Ansätzen werden diese Kandidaten jedoch nur als gültig betrachtet, falls diese ebenfalls im Rezensionstext vorkommen. Eine Berücksichtigung von impliziten Aspekten findet bei diesem Ansatz nicht statt. Für die Gruppierung wird ein Ansatz aus dem Bereich des unüberwachten Lernens (engl. *unsupervised learning*) verwendet, welcher auf dem Topicmodell Latent Dirichlet Allocation (LDA) basiert. Kategorien können hier nicht manuell festgelegt werden, sondern nur die Anzahl der Cluster (Topics), in welche die Produktraspekte aufgeteilt werden sollen. Eine Untersuchung von Chen et al. hat gezeigt, dass der Einsatz von Topicmodellen wie LDA für kürzere Texte (z.B. Microblogging-Nachrichten) eher ungeeignet ist [28, S. 2418].

Der Ansatz von Zhai et al. beschäftigt sich nur mit der Gruppierung von Produktraspekten und setzt eine korrekte Extraktion von Aspektbezeichnungen durch bestehende Systeme voraus [219, S. 1273]. Zum Einsatz kommt ein Semi-Supervised Ansatz basierend auf Naïve Bayes und Expectation Maximization (EM). Innerhalb der Untersuchung werden vollständige Rezensionstexte (Format F_D) betrachtet. Einem Teil dieser Rezensionen wird manuell eine Produktraspektgruppe zugewiesen, während der andere Teil der Rezensionen ungekennzeichnet bleibt. Danach wird ein Klassifikator entwickelt, welcher bestehend auf den gekennzeichneten Datensätzen die nicht gekennzeichneten Datensätze klassifiziert. Für jedes Wort des Datensatzes wird dessen jeweiliger Kontext ermittelt, d.h., abhängig von einer definierten Fenstergröße n , werden die n Wörter links und rechts des entsprechenden Wortes als Kontext vermerkt. Die Ergebnisse des Ansatzes sind allerdings dahin gehend zu relativieren, dass für eine Anwendung des Ansatzes eine völlig korrekte Identifikation der Aspektbezeichnungen vorausgesetzt wird, wovon in der Praxis jedoch nicht ausgegangen werden kann.

Einen etwas unterschiedlichen Anwendungsfall zu den oben genannten Arbeiten liefert die Arbeit von Lu et al. Hier werden drei Clustering-Ansätze untersucht, welche die

Tabelle 7.1: Abgrenzung der verwandten Arbeiten aus den Bereichen Identifikation und Gruppierung von Aspekten

Ansatz	Format	Implizite Aspekte	Gruppierung
Carenini et al. [25]	F_D	nein	supervised
Zhai et al. [219]	F_D	nein	supervised
Popescu und Etzioni [159]	F_D	ja	n/a
Hu und Liu [73, 74]	F_D	nein	n/a
Guo et al. [56]	F_{PKD}	nein	unsupervised
Lu et al. [125]	F_{PK}	nein	unsupervised
Hu und Liu [75]	F_{PK}	ja	n/a
Liu et al. [119]	F_{PK}	ja	supervised
<i>Ansatz dieser Arbeit</i>	F_{PK}	ja	supervised

Aspekte bei der Bewertung von Verkäufern auf der Auktionsplattform eBay¹ aggregieren [125]. Bei den untersuchten Inhalten dieses Ansatzes handelt es sich um sehr kurze Benutzerkommentare und somit um eine Anwendung von Rezensionsformat F_{PK} . Allerdings unterscheidet sich die Untersuchung von Lu et al. darin, dass die Eigenschaften eines Verkäufers bewertet werden und nicht die Merkmale von Produkten. Für die Gruppierung wird ein unüberwachtes Clustering verwendet, wobei die Anzahl der resultierenden Cluster auf acht begrenzt wurde. Diese Restriktion ist jedoch für Produktaspekte nicht sinnvoll anwendbar, da Produktgruppen häufig deutlich mehr als acht verschiedene Produktaspekte umfassen.

Abschließend sollen die oben genannten Arbeiten tabellarisch zusammengefasst und zum Ansatz dieser Arbeit abgegrenzt werden. Tabelle 7.1 verdeutlicht hierzu die Ansätze, das jeweils verwendete Rezensionsformat, die Berücksichtigung von impliziten Aspekten sowie die Methode zur Gruppierung der Aspekte (falls anwendbar). Arbeiten die zur Gruppierung sowohl einen überwachten (supervised) als auch einen nicht-überwachten (unsupervised) Ansatz verwenden, werden in die Kategorie *supervised* eingeordnet. Arbeiten die nur eine Identifikation (und keine Gruppierung) von Aspekten vornehmen, sind in der Spalte „Gruppierung“ mit „n/a“ gekennzeichnet. Wie die letzte Zeile aus Tabelle 7.1 zeigt, liegt der Fokus dieser Arbeit in der Betrachtung des Rezensionsformats F_{PK} und der Identifikation von impliziten Aspekten unter der Anwendung eines supervised-Ansatzes zur Gruppierung der Aspekte. Diese Arbeit ist damit direkt mit der Arbeit von Liu et al. [119] vergleichbar.

¹<http://www.ebay.com> (besucht am 11.02.2015)

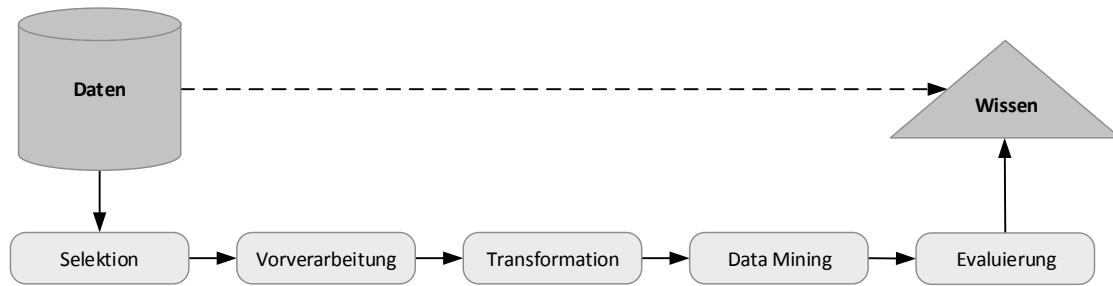


Abbildung 7.4: Schritte des KDD-Prozesses (in Anlehnung an [47, S. 41])

7.3 Klassifikationsmodell

Nach der Einordnung und Abgrenzung des Problems in den vorigen Abschnitten soll nun ein Klassifikationsmodell vorgestellt werden, um die erstellten Rated Tags in semi-automatischer Weise zu vereinheitlichen. Zur Verdeutlichung wird in Abschnitt 7.3.1 zunächst eine Darstellung des Klassifikationsproblems vorgenommen. Abschnitt 7.3.2 zeigt Ansätze zur Vorverarbeitung der Daten, um die Qualität der Algorithmen zu verbessern. Die weiteren Abschnitte 7.3.3 bis 7.3.5 betrachten dann die entwickelten Klassifikationsalgorithmen.

7.3.1 Modellübersicht

Das vorliegende Problem kann dem Bereich *Knowledge discovery in databases (KDD)* zugeordnet werden, da versucht wird aus der Analyse der Daten neues Wissen abzuleiten. Das Problem lässt sich daher auf den abstrakten KDD-Prozess von Fayyad et al. abbilden (siehe Abbildung 7.4). Der Prozess beschreibt die Selektion der Daten sowie deren Vorverarbeitung und Transformation. Im vorletzten Prozessschritt „Data Mining“ kommen dann Klassifikations- bzw. Clusteringverfahren zum Einsatz, welche in einem abschließenden Prozessschritt evaluiert werden.

Wie bereits in Abschnitt 7.2.1 erläutert wurde, ist für die semi-automatische Vereinheitlichung der Rated Tags eine Benutzerbeteiligung erforderlich, d.h., für den Prozessschritt „Data Mining“ werden Methoden des überwachten Lernens (engl. *supervised learning*) eingesetzt. Es handelt sich also um den Einsatz von Klassifikationsverfahren und nicht um Clusteringverfahren, wie es typischerweise im Bereich des unüberwachten Lernens (engl. *unsupervised learning*) der Fall ist.

Das Ziel ist die Klassifikation bzw. Kategorisierung von Dokumenten. Es existiert also eine Dokumentenmenge $D = \{d_1, d_2, \dots, d_{|D|}\}$, wobei jedes Dokument d_i genau einem Rated Tag entspricht. Ein Dokument d_i besteht also wiederum aus einem oder mehreren Wörtern bzw. Begriffen (engl. *terms*) $T_i = \{t_{i1}, t_{i2}, \dots, t_{i|T_i|}\}$, wobei aufgrund der Charakteristik von Rated Tags, $|T_i|$ meist wesentlich kleiner als 10 ist. Ein weiterer Bestandteil eines Dokuments d_i ist die Bewertungskomponente b_i , welche Werte innerhalb eines Intervalls $[I_S, I_E] = \{I_S \leq b_i \leq I_E, b_i \in \mathbb{N}\}$ annehmen kann. Abhängig von dem verwendeten Bewertungsintervall, welches typischerweise von 1 bis 5 reicht, gilt also für

jedes Element $b_i \in [I_S, I_E]$. Damit kann jedes Dokument d_i als geordnetes Paar dargestellt werden mit $d_i = (T_i, b_i)$.

Die Betreiber des Rated Tags-Systems definieren für jede Produktgruppe zunächst eine Menge an Klassen mit $C = \{c_1, c_2, \dots, c_{|C|}\}$, in welche später alle Dokumente eingeordnet werden sollen. Eine solche Klassifikation ist natürlich nur sinnvoll für $|C| \geq 2$. Im Gegensatz zur binären Klassifikation (engl. *binary classification*) mit $|C| = 2$ können bei der vorliegenden Problemstellung mehr als zwei Klassen definiert werden, d.h., es handelt sich um eine Mehrklassen-Klassifikation (engl. *multiclass classification*). Ein Beispiel für eine binäre Klassifikation wäre ein Spamfilter für E-Mails, welcher eingehende E-Mails entweder als Spam oder Nicht-Spam klassifiziert.

Nach der Erzeugung der Klassen C wird jedes Element l einer Teilmenge $L \subseteq D$ mit einer Klasse c gekennzeichnet, d.h. $(l, c) \in L \times C$. An dieser Stelle sei erwähnt, dass es sich bei dem vorliegenden Problem um eine *Single-Label*-Klassifikation handelt, d.h., jedem Dokument wird genau eine Klasse zugewiesen [173, S. 3]. Die Teilmenge der gekennzeichneten (engl. *labeled*) Dokumente L wird auch als *Trainingsset* bezeichnet, welches für das Training einer Klassifikationsfunktion γ (auch Klassifikationsmodell oder kurz Klassifikator genannt) verwendet wird. Die Aufgabe eines Klassifikators ist es dann, jedem Element der nicht gekennzeichneten (engl. *unlabeled*) Dokumentenmenge $U = D \setminus L$ jeweils eine passende Klasse aus C zuzuweisen, d.h. $\gamma : U \rightarrow C$. Für die in den Abschnitten 7.3.3 bis 7.3.5 dargestellten Klassifikationsalgorithmen wird allerdings nicht die wahrscheinlichste Klasse, sondern ein Wahrscheinlichkeitsvektor W über alle Klassen benötigt, weshalb die Klassifikationsfunktion wie folgt formuliert wird: $\gamma : U \rightarrow W$.

Das typische Vorgehen für die Erstellung eines Klassifikators aus dem Bereich des überwachten Lernens ist in Abbildung 7.5 dargestellt. Die Abbildung zeigt die Zuweisung von Klassen an eine Dokumentenmenge. Diese Menge wird dann in Trainingsdaten (Trainingsset) und Testdaten (Testset) unterteilt. Dem zu entwickelnden Lernalgorithmus stehen die Trainingsdaten zur Verfügung, d.h., zu jeder Instanz in den Trainingsdaten ist die dafür zugewiesene Klasse ersichtlich. Der Lernalgorithmus versucht nun die Zusammenhänge aus den Trainingsdaten zu analysieren und weitestgehend zu verallgemeinern. Das Ergebnis des Lernalgorithmus ist eine Klassifikationsfunktion, welche basierend auf dem erlernten Wissen in der Lage ist für eine nicht-klassifizierte Instanz eine Klasse vorherzusagen. Hierzu verarbeitet der Klassifikator die in Abbildung 7.5 dargestellten Testdaten. Die Testdaten enthalten zwar bereits die korrekten Klassen, allerdings wird diese Information dem Klassifikator vorenthalten. Die Güte des Klassifikators wird dann evaluiert, indem die vorhergesagte Klasse für eine Testdaten-Instanz mit der tatsächlichen Klasse verglichen wird.

Für oben genannte Problemstellung kommen unter anderem die gängigen Algorithmen der überwachten Textklassifikation infrage. Aggarwal und Zhai nennen hierzu Naïve Bayes, Support Vector Machine (SVM), Entscheidungsbaumklassifikatoren und einige mehr [2, S. 165 f.]. Der Naïve Bayes-Ansatz gilt als performant, einfach zu implementieren und relativ effizient [163, S. 616]. Laut Liu et al. hat sich jedoch SVM zum populärsten Lernalgorithmus für Textklassifikation entwickelt [122, S. 174]. Ein Nachteil von SVM ist allerdings die hohe Implementierungskomplexität im Vergleich zu Naïve Bayes. Rennie et al. zeigen, dass ein Naïve Bayes-Ansatz bei entsprechender Vorverarbeitung der Daten eine

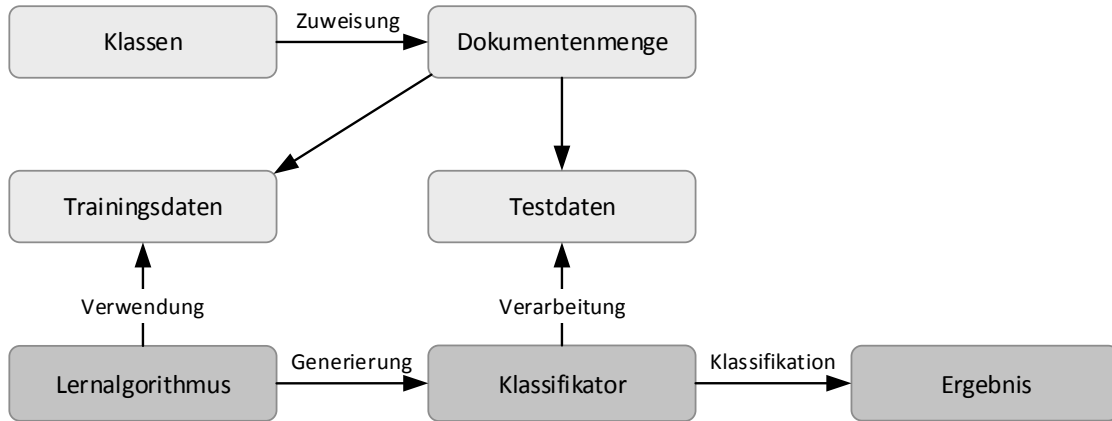


Abbildung 7.5: Prozessübersicht zur Erstellung eines Klassifikators

nahezu ähnliche Klassifikationsgüte erreichen kann wie ein State-of-the-art SVM-Ansatz [163, S. 616].

In der späteren Evaluierungsphase werden sowohl der Naïve Bayes-Ansatz auch als der SVM-Ansatz getestet. Beim Naïve Bayes-Ansatz werden zudem noch das Multinomial- und das Bernoulli- Dokumentenmodell unterschieden. Jeder dieser drei Hauptklassifikatoren wird zusammen mit den anderen String-basierten Klassifikatoren getestet, um den am besten geeigneten Hauptklassifikator zu ermitteln.

Für Instanzen, welche typischerweise nicht oder nur schwer von den Hauptklassifikatoren erfasst werden können, werden String-basierte Algorithmen implementiert. Der Lernalgorithmus des Hauptklassifikators ergibt zusammen mit den String-basierten Algorithmen also eine Menge an Basislernalgorithmen $A = \{a_1, a_2, \dots, a_{|A|}\}$, welche durch die Anwendung der Trainingsdaten in eine Klassifikatorenmenge $K = \{k_1, k_2, \dots, k_{|A|}\}$ überführt werden. Jeder Klassifikator k_i definiert eine Funktion γ , um eine nicht gekennzeichnete Instanz u in einen Wahrscheinlichkeitsvektor $W_u = (w_{uc_1}, w_{uc_2}, \dots, w_{uc_{|C|}})$ der Länge $|C|$ zu überführen. Der Vektor enthält für jede Klasse c_i die Wahrscheinlichkeit auf einem Intervall von 0 bis 1, dass die nicht gekennzeichnete Instanz u zu dieser Klasse gehört. Somit lautet die Funktion γ :

$$\gamma : U \rightarrow W, u \mapsto (w_{uc_1}, w_{uc_2}, \dots, w_{uc_{|C|}}), w_{uc_i} \in \mathbb{R}, w_{uc_i} \in [0, 1] \quad (7.3)$$

Da es sich wie oben erwähnt um mehrere Klassifikatoren handelt, wird eine Aggregation der Klassifikationsergebnisse benötigt. Hierzu wird ein sogenannter *Ensemble-Klassifikator* entwickelt, welcher die Wahrscheinlichkeitsvektoren der einzelnen Klassifikatoren aggregiert und dann darüber entscheidet, welche Klasse die höchste Wahrscheinlichkeit für die zu klassifizierende Instanz besitzt.

Ein bislang nicht erwähnter Prozessschritt ist die Vorverarbeitung der Daten. Dies ist im vorliegenden Problem besonders wichtig, da es sich um Benutzer-generierte Daten handelt, welche üblicherweise nicht der Qualität redaktionell geprüfter Inhalte, vor allem in Hinsicht auf Satzbau und Rechtschreibung, entsprechen. Es werden verschiedene

Verfahren eingesetzt um die Daten zu bereinigen und die Dimensionalität der Daten zu reduzieren, z.B. durch das Entfernen von nicht relevanten Wörtern (Stoppwörtern).

Eine zusammenfassende Darstellung der Bestandteile und Prozessschritte innerhalb des Klassifikationsmodells ist Abbildung 7.6 zu entnehmen. In den nachfolgenden Abschnitten findet eine detaillierte Betrachtung der Bestandteile statt. Zunächst werden die Verfahren zur Vorverarbeitung der Daten in Abschnitt 7.3.2 näher erläutert. In Abschnitt 7.3.3 und 7.3.4 werden die konzipierten Basis-Lernalgorithmen näher betrachtet. Abschnitt 7.3.5 erläutert dann die Konzeptionierung des Ensemble-Klassifikators und die Aggregation der einzelnen Basis-Klassifikatoren.

7.3.2 Vorverarbeitung der Daten

Da es sich bei Rated Tags um sehr kurze Texte handelt, ist eine Vorverarbeitung und Normalisierung der Daten sehr wichtig, da ohnehin nur ein geringer Kontext existiert. Für das Modell werden die folgenden Schritte vorgeschlagen, welche nachfolgend näher erläutert werden:

- Anpassung des Textes auf Zeichenebene
- POS-Tagging
- Tokenisierung
- Stoppwortentfernung
- Normalisierung und Lemmatisierung
- Stammformreduktion

Der erste Schritt der Datenvorverarbeitung ist das Entfernen von Sonderzeichen, welche für die Textklassifikation typischerweise keine Bedeutung haben. Als Sonderzeichen werden alle nicht alphanumerischen Zeichen, d.h., alle Zeichen mit Ausnahme von Buchstaben und Ziffern, angesehen. Diese Zeichen werden jeweils durch ein Leerzeichen ersetzt. Danach werden alle nicht direkt von Buchstaben umgebenden numerischen Zeichen entfernt, z.B. „300“, jedoch nicht „3d“. Noch vor dem Entfernen der Sonderzeichen wird innerhalb der Tags nach Währungsbeträgen gesucht, z.B. „\$100“ oder „300 €“. Da der genaue Betrag keine Rolle spielt, werden die gefundenen Textblöcke durch „\$1“ ersetzt, um den Feature-Vektor zu reduzieren. Das gleiche Normalisierungsverfahren wird für Zeiteinheiten gewählt, d.h., alle zeitlichen Angaben wie „Month“, „Week“ etc. werden auf „Month“ normalisiert. Abschließend werden alle Buchstaben des Textes in Kleinbuchstaben (engl. *lower case*) konvertiert.

Für eine spätere Verwendung einer lexikalischen Datenbank wie beispielsweise WordNet, ist die Bestimmung der Wortarten für die im Text befindlichen Wörter notwendig. Für diesen Zweck kann ein POS-Tagger eingesetzt werden. Der Stanford Part-of-Speech Tagger¹ ist hier einer der verbreitetsten seiner Art für englischsprachige Texte. Er wird ebenfalls im Rahmen der späteren Evaluierung verwendet.

¹<http://nlp.stanford.edu/software/tagger.shtml> (besucht am 15.03.2015)

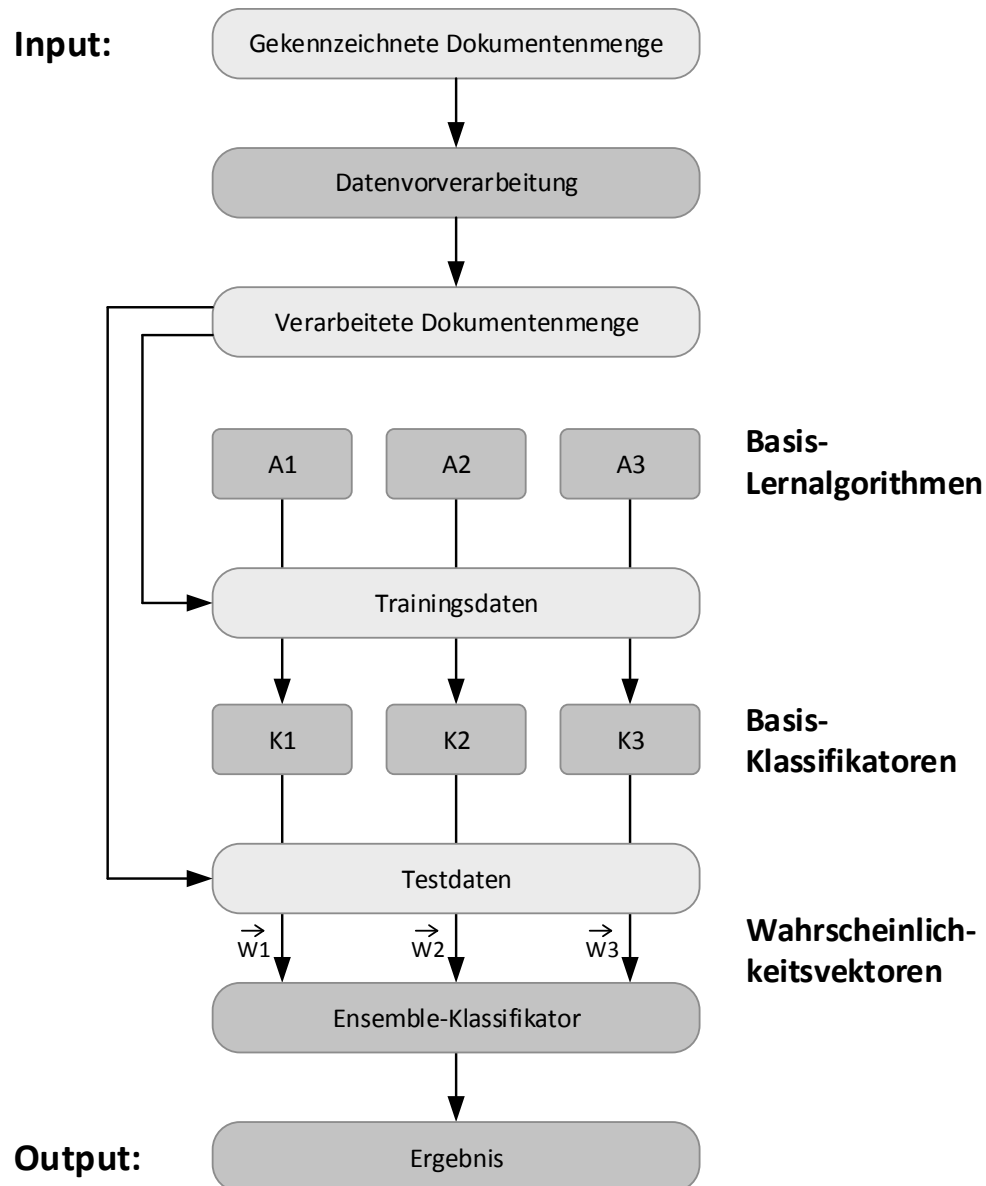


Abbildung 7.6: Bestandteile und Prozessschritte innerhalb des Klassifikationsmodells

Im nächsten Schritt der Vorverarbeitung wird eine Tokenisierung des Textes, d.h. eine Zerlegung in einzelne Wörter, vorgenommen. Dabei werden ein oder mehrere Leerzeichen als Begrenzung zwischen den Wörtern angesehen. Dadurch entsteht eine Multimenge an Wörtern (engl. *bag of words*). Es wird allerdings davon ausgegangen, dass innerhalb eines Tags ein Wort meist nur einmal vorkommt.

Eine weitere wichtige Aufgabe der Datenvorverarbeitung ist das Entfernen von sogenannten Stoppwörtern. Als Stoppwort wird ein Wort bezeichnet, welches häufig in Texten vorkommt, jedoch keine Relevanz hinsichtlich der Klassifikation besitzt. Beispielhafte Stoppwörter der englischen Sprache sind „the“, „of“, „is“ und viele mehr. Das Entfernen von Stoppwörtern ist aus zweierlei Gründen sinnvoll. Zum einen verbessert es die Klassifikationsergebnisse, da die nicht relevanten Wörter nicht in den Klassifikationsalgorithmus einfließen. Zum anderen wird durch das Entfernen der Stoppwörter auch die Länge des Feature-Vektors reduziert, was zu einer Verbesserung der Laufzeitleistung führt. Ein Beispiel hierzu ist der Satz „The picture of the TV is good“. Ohne eine Entfernung der Stoppwörter würde sich ein Feature-Vektor der Länge 6 ergeben (*'the', 'picture', 'of', 'TV', 'is', 'good'*). Ohne Stoppwörter beträgt die Länge des Vektors hingegen nur 3 (*'picture', 'TV', 'good'*).

Da keine standardisierten Stoppwortlisten existieren, wurden diverse Listen^{1,2} zusammengetragen und aggregiert. Zusätzlich zu „normalen“ Stoppwörtern, existieren in Bezug auf die Klassifikation weitere Wörter, welche als nicht relevant eingestuft werden können. Hierbei handelt es sich um neutrale Adjektive, welche keinen Mehrwert hinsichtlich der Klassifikation aufweisen. Beispiele sind die Adjektive „good“ oder „bad“, welche für unterschiedliche Merkmale verwendet werden können, z.B. „good sound“ und „good picture“. Solche Adjektive werden im Rahmen der Klassifikation ebenfalls als Stoppwörter betrachtet und entsprechend entfernt. Im Gegensatz dazu existieren Adjektive, welche implizit auf Produktaspekte schließen lassen. So lässt das Adjektiv „big“ aus dem Tag „big screen“ auf den Produktaspekt „Größe“ schließen. Solche relevanten Tags werden nicht entfernt.

Für das vorliegende Klassifikationsproblem spielt die Unterscheidung von positiven oder negativen Meinungen keine Rolle, da diese bereits durch die Bewertungskomponente der Rated Tags festgelegt werden. Daher wird im Rahmen der Normalisierung vorgeschlagen, alle vorkommenden negativen Präfixe zu entfernen. Beispiele negativer Präfixe sind „in-“ und „un-“, welche beispielsweise in den Wörtern „inaccurate“ und „unreliable“ enthalten sind. Die Entfernung dieser Präfixe wird vorgeschlagen, da diese vor allem bei Adjektiven auftreten, welche wiederum häufig Produkteigenschaften umschreiben. Das Ziel ist die Vereinheitlichung der Wörter, da ein Klassifikator beispielsweise die Wörter „inaccurate“ und „accurate“ als unterschiedliche Features interpretiert, obwohl diese den gleichen Aspekt adressieren. Um die potenziellen Kandidaten für die Entfernung eines Präfixes zu ermitteln, wird folgender regulärer Ausdruck verwendet: $\wedge(un|il|im|ir|in|dis|mis|non)(\wedge w+)\$$

¹<http://www.ranks.nl/stopwords> (besucht am 15.03.2015)

²<http://xpo6.com/list-of-english-stop-words/> (besucht am 15.03.2015)

Für alle Wörter die dem oben genannten regulären Ausdruck entsprechen, werden mittels WordNet die zugehörigen Antonyme ermittelt. Existiert ein Antonym das dem Wort ohne Präfix entspricht, so wird der Präfix des jeweiligen Wortes entfernt. Dies führt mitunter zu einer Reduktion des Feature-Vektors.

Ein weiterer Schritt zur Normalisierung der Daten erfolgt durch die Lemmatisierung. Dabei wird die lexikalische Datenbank WordNet verwendet, um die Grundform eines Wortes zu ermitteln. Für diese Ermittlung werden unter anderem die oben genannten POS-Informationen der jeweiligen Wörter herangezogen. Die Vereinheitlichung der Wörter durch die Lemmatisierung dient ebenfalls der Verringerung des Feature-Vektors. So werden beispielsweise die Wörter „better“ und „feet“ durch eine Lemmatisierung auf die Grundformen „good“ und „foot“ reduziert.

Als abschließender Prozessschritt der Datenvorverarbeitung findet eine Stammformreduktion (engl. *stemming*) aller vorkommenden Wörter statt. Stemming unterscheidet sich zur Lemmatisierung darin, dass Stemming durch einen heuristischen Prozess versucht den Wortstamm zu bilden, während bei der Lemmatisierung typischerweise auf Wörterbücher (im oben beschriebenen Fall auf WordNet) zurückgegriffen wird [127, S. 32]. Der wohl am Häufigsten eingesetzte Stemming-Algorithmus für englische Texte ist der Porter-Stemmer¹ von Martin Porter [160], welcher ebenfalls in der späteren Evaluierung verwendet wird.

Es muss berücksichtigt werden, dass Wörter durch den Einsatz von Stemming teilweise drastisch gekürzt werden und möglicherweise nicht mehr in einem Wörterbuch auffindbar sind. So würde der Porter-Stemmer das Wort „computing“ beispielsweise auf den Wortstamm „comput“ reduzieren, welcher kein gültiges Wort mehr darstellt. Daher besteht durch Stemming allerdings auch die Gefahr, die Klassifikation zu verschlechtern. Ein Negativbeispiel zeigen Metzler et al. anhand der beiden Textausschnitte „marine vegetation“ und „marinated vegetables“ [132, S. 3]. Ein Porter-Stemmer würde beide Texte auf „marin veget“ reduzieren, wodurch ein Klassifikator beide Texte als gleichwertig betrachten würde. Metzler et al. merken jedoch an, dass die Vorteile von Stemming überwiegen und empfehlen daher den Einsatz [132, S. 3]. Laut Liu existiert trotz vieler Untersuchungen kein endgültiger Nachweis, dass Stemming das Klassifikationsergebnis tatsächlich verbessert [118, S. 228]. In Abschnitt 7.4 wird im Rahmen der Evaluierung unter anderem untersucht, ob und wie Stemming das Klassifikationsergebnis für die vorliegende Problemstellung beeinflusst.

7.3.3 Hauptklassifikatoren

Allgemeines

Für die Generierung des Klassifikators werden in der Trainingsphase einige Informationen benötigt. Zunächst muss das *Vokabular* V aller Trainingsdokumente bestimmt werden. Die Vokabularmenge ergibt sich durch die Vereinigung aller Begriffe T der Trainingsdokumente L (siehe Gleichung 7.4). Für jeden Term t_{ij} eines Dokuments L_i gilt folglich $t_{ij} \in V$. In

¹<http://tartarus.org/martin/PorterStemmer/> (besucht am 15.03.2015)

der späteren Testphase ist es jedoch durchaus möglich, dass ein Term t_{ij} eines Dokuments U_i nicht im Vokabular V enthalten ist, d.h., der Fall $t_{ij} \notin V$ kann in der Testphase eintreten.

$$V = \bigcup_{i=1}^{|L|} T_i \quad (7.4)$$

Basierend auf dem Vokabular V lässt sich ein binärer Wort-Vektor Z der Länge $|V|$ abbilden. Jeder in Dokument L_i vorkommende Term t_{ij} erhält im Wort-Vektor Z_i den Wert 1, alle anderen Elemente den Wert 0, d.h. $z_{ij} \in \{0, 1\}$. Zusätzlich wird für die Klassifikation die Bewertungskomponente b_i eines jeden Dokumentes L_i herangezogen, da davon ausgegangen wird, dass die Bewertung eines Rated Tags zur Bestimmung der korrekten Klasse beitragen kann. Existieren beispielsweise zwei Klassen „Klang“ und „Preis“, wovon Erstere überwiegend positiv und Letztere überwiegend negativ bewertet wurde, dann besitzt ein noch nicht klassifiziertes, eher negativ bewertetes Dokument eine höhere Wahrscheinlichkeit der Klasse „Preis“ anzugehören.

Um diesen Sachverhalt abzubilden, wird die Bewertungskomponente b_i ebenfalls in einen binären Vektor umgewandelt. Die Länge des Vektors könnte 2 betragen, falls nur eine Unterteilung in positiv und negativ vorgenommen wird. Hier wäre jedoch eine Unterscheidung von neutralen Bewertungen, z.B. 4 von 10 und 6 von 10, problematisch, da Ersteres als negativ und Letzteres als positiv gewertet würde. Um auch eher neutrale Bewertungen in einem separaten Intervall zusammenzufassen, wurde für den Vektor die Länge 3 gewählt, d.h., eine Bewertung b_i wird entweder als negativ, neutral oder positiv eingestuft. Die Funktion $\theta(b_i)$ aus Gleichung 7.5 zeigt die Überführung einer Bewertungskomponente b_i in einen binären Vektor der Länge 3. Der Parameter I_E steht dabei für den maximalen Wert des Bewertungsintervalls.

$$\theta(b_i, I_E) = \begin{cases} (1, 0, 0) & \text{falls } 0 \leq \frac{b_i}{I_E} < \frac{1}{3} \\ (0, 1, 0) & \text{falls } \frac{1}{3} \leq \frac{b_i}{I_E} < \frac{2}{3} \\ (0, 0, 1) & \text{sonst} \end{cases} \quad (7.5)$$

Da nun sowohl die Begriffe eines Rated Tags als auch dessen Bewertung in ein binäres Vektorformat konvertiert wurden, lassen sich diese beiden Vektoren zusammenfügen und ergeben so für jedes Dokument d_i den *Feature-Vektor* F_i aus Gleichung 7.6. Im Gegensatz zu traditionellen Ansätzen der Textklassifikation, bei welchen der Feature-Vektor F_i dem Wort-Vektor Z_i entspricht, werden hier ebenfalls die Bewertungskomponenten b_i berücksichtigt.

$$F_i = Z_i \oplus \theta(b_i, I_E) \quad (7.6)$$

Zur beispielhaften Verdeutlichung des obigen Sachverhalts wird angenommen, dass die Trainingsdaten aus den Rated Tags-Dokumenten l_1 und l_2 aus Gleichung 7.7 bestehen. Für die Bewertungskomponente b_i und dessen möglichen Wertebereich gilt $1 \leq b_i \leq 5, b_i \in \mathbb{N}$. Eine Vereinigung der in den Rated Tags enthaltenen Begriffsmengen T_1 und T_2 ergibt das Vokabular V aus Gleichung 7.8. Basierend auf dem Vokabular, lassen sich für die beiden Dokumente die binären Wort-Vektoren Z_1 und Z_2 der Länge $|V|$ erzeugen (siehe

Tabelle 7.2: Beispielhafte Feature-Vektoren für die in Gleichung 7.7 gezeigten Rated Tags-Dokumente l_1 und l_2

F_i	Z_i						$\theta(b_i, 5)$		
	the	picture	is	best	sound	awesome			
F_1	1	1	1	1	0	0	0	0	1
F_2	1	0	1	0	1	1	0	0	1

Tabelle 7.2). Die Bewertungskomponenten b_i werden gemäß der Funktion $\theta(b_i, I_E)$ aus Gleichung 7.5 mit $I_E = 5$ jeweils in einen binären Vektor überführt. Durch die Verkettung der beiden binären Vektoren ergeben sich schließlich die in Tabelle 7.2 dargestellten Feature-Vektoren F_1 und F_2 .

$$\begin{aligned} l_1 &= ('the\ picture\ is\ the\ best', 5) \\ l_2 &= ('the\ sound\ is\ awesome', 4) \end{aligned} \quad (7.7)$$

$$V = \{the, picture, is, best, sound, awesome\} \quad (7.8)$$

Die in Tabelle 7.2 dargestellten Wort-Vektoren Z_i zeigen zum einen, dass auch ein mehrfach vorkommendes Wort (z.B. „the“ in l_1) im Rahmen dieses Ansatzes maximal den Wert 1 erhält, d.h., es wird immer ein binärer Feature-Vektor verwendet. Es wird davon ausgegangen, dass die Worthäufigkeit innerhalb eines Tags für die Klassifikation keinen Mehrwert besitzt. Ist ein Wort aus dem Vokabular V nicht innerhalb eines Wort-Vektors vorhanden, so wird dessen Wert auf 0 gesetzt. Zum anderen verdeutlicht die Tabelle, dass es für die Trainingsphase notwendig ist, nur bestimmte Features zu selektieren. Im Beispiel aus Tabelle 7.2 werden die Wörter „the“ und „is“ miteinbezogen. Dadurch erhöht sich (fälschlicherweise) die Ähnlichkeit zwischen den Wort-Vektoren Z_1 und Z_2 . Solche sogenannten Stoppwörter werden daher im Rahmen der Datenvorverarbeitung entsprechend entfernt (siehe Abschnitt 7.3.2).

Naïve Bayes Klassifikator

Ein Naïve Bayes-Klassifikator erzeugt für eine bestimmte Instanz eine Wahrscheinlichkeitsverteilung für alle infrage kommenden Klassen. Er beruht auf dem *Satz von Bayes* (*Bayes Theorem*) zur Berechnung bedingter Wahrscheinlichkeiten. Übertragen auf die Klassifikation von Dokumenten kann das Bayes Theorem wie folgt formuliert werden:

$$P(c|d) = \frac{P(d|c) \cdot P(c)}{P(d)} \quad (7.9)$$

Die Wahrscheinlichkeit $P(c)$ entspricht der a priori-Wahrscheinlichkeit für das Vorkommen der Klasse c in einer Dokumentenmenge D . Die Wahrscheinlichkeit $P(d)$ steht für die a priori-Wahrscheinlichkeit, dass ein Dokument d auftritt. Die bedingte Wahrscheinlichkeit $P(d|c)$ entspricht der Wahrscheinlichkeit, dass ein Dokument d einer

Tabelle 7.3: Unterschiede zwischen einem Bernoulli und Multinomial Naïve Bayes-Dokumentenmodell (in Anlehnung an [127, S. 268])

	Bernoulli	Multinomial
Dokumenten- abbildung	Binärer Vektor für das (Nicht-) Vorkommen von Wörtern	Vektor von Worthäufigkeiten
Berücksichtigung Worthäufigkeiten	Nein	Ja
Stärken	Kürzere Dokumente	Längere Dokumente
Auswirkung nicht- vorkommender Wörter	Beeinflusst Wahrscheinlichkeit	Kein Einfluss

Klasse c angehört. Die gesuchte Wahrscheinlichkeit $P(c|d)$ steht schließlich für die a posteriori-Wahrscheinlichkeit, dass ein Dokument d zu einer Klasse c gehört. Da die Wahrscheinlichkeit für das Vorkommen eines Dokuments für alle Klassen identisch ist, kann der Nenner $P(d)$ aus Gleichung 7.9 entfernt werden [127, S. 265], wodurch sich folgende Proportionalität (\propto) ergibt:

$$P(c|d) \propto P(d|c) \cdot P(c) \quad (7.10)$$

Für die Berechnung der Wahrscheinlichkeiten ist relevant, wie ein Dokument d im Rahmen des Naïve Bayes-Ansatzes interpretiert wird. Hierzu gibt es zwei verschiedene Naïve Bayes-Dokumentenmodelle, nämlich das (*multivariate*) *Bernoulli-Dokumentenmodell* und das *Multinomial-Dokumentenmodell*. Für die Klassifikation von längeren Texten (wie beispielsweise auch Rezensionen) wird typischerweise ein Multinomial Naïve Bayes verwendet, da dieser unter anderem das mehrfache Vorkommen von Wörtern berücksichtigt. Das Naïve Bayes Bernoulli-Modell berücksichtigt hingegen nur das Vorhandensein, aber auch das Nicht-Vorhandensein von Wörtern und eignet sich besonders für kürzere Dokumente [127, S. 268]. Da es sich bei den Rated Tags um sehr kurze Texte handelt, wurde für die Klassifikation das Naïve Bayes Bernoulli-Modell gewählt. Auf eine nähere Erläuterung des multinomialen Naïve Bayes-Ansatzes wird daher an dieser Stelle verzichtet. Alle nachfolgenden Beschreibungen beziehen sich deshalb, sofern nicht anders angegeben, auf das Bernoulli-Dokumentenmodell. Die wichtigsten Unterschiede beider Modelle sind in Tabelle 7.3 kurz zusammengefasst.

Die obigen Ausführungen haben gezeigt, wie Rated Tags-Dokumente in binäre Feature-Vektoren überführt werden können. Für jedes Feature f_{ij} eines Dokuments d_i gilt daher $f_{ij} \in \{0, 1\}$. Da das Bernoulli-Dokumentenmodell sowohl das Vorkommen als auch das Fehlen eines Features berücksichtigt, kann die Wahrscheinlichkeit eines Features f_{ij} bei gegebener Klasse c_m gemäß Gleichung 7.11 ermittelt werden. F_{c_m} steht dabei für alle Feature-Vektoren, welche mit der Klasse c_m gekennzeichnet sind.

$$P(f_{ij} | c_m) = f_{ij} \cdot \frac{\sum_{y=1}^{|Fc_m|} f_{yj}}{|Fc_m|} + (1 - f_{ij}) \cdot \left(1 - \frac{\sum_{y=1}^{|Fc_m|} f_{yj}}{|Fc_m|}\right) \quad (7.11)$$

Für die weitere Betrachtung des Lernalgorithmus sollen zunächst die bislang unberücksichtigten Eigenschaften von Naïve Bayes erläutert werden. Und zwar trifft der Naïve Bayes-Algorithmus hinsichtlich der Elemente innerhalb des Feature-Vektors die folgenden zwei wichtigen Annahmen [2, S. 182]:

1. Die Reihenfolge innerhalb eines Wort-Vektors spielt keine Rolle (Bag of words (BOW)-Annahme)
2. Für eine gegebene Klasse sind die im Feature-Vektor enthaltenen Elemente untereinander unabhängig (Unabhängigkeitsannahme)

Die erste Annahme wird auch häufig als BOW-Annahme bezeichnet, da der Naïve Bayes-Ansatz den Wort-Vektor als Multimenge (engl. *bag*), d.h., als Menge in der Elemente mehrfach vorkommen dürfen, betrachtet. Wie bereits aus Tabelle 7.2 hervorging, wird die Reihenfolge der ursprünglichen Begriffe im Wort-Vektor nicht berücksichtigt. Obwohl die oben genannten Annahmen in der Praxis meist nicht erfüllt sind, erzielt Naïve Bayes gute Klassifikationsergebnisse [131, S. 41]. Domingos und Pazzani zeigen ferner, dass die Unabhängigkeitsannahme für Attribute nicht erfüllt sein muss, damit der Naïve Bayes Algorithmus optimal arbeitet [43, S. 111].

Durch die oben genannte Unabhängigkeitsannahme zwischen den einzelnen Features lässt sich die bedingte Wahrscheinlichkeit $P(F_i | c_k)$ als Produkt der einzelnen Wahrscheinlichkeiten ausdrücken.

$$P(F_i | c_k) = \prod_{j=1}^{|F_i|} P(f_{ij} | c_k) \quad (7.12)$$

Die a priori-Wahrscheinlichkeit für das Vorkommen einer Klasse c lässt sich wie folgt ermitteln:

$$P(c) = \frac{|D_c|}{|D|} \quad (7.13)$$

D_c stellt dabei die Menge aller Dokumente dar, welche der Klasse c zugewiesen wurden. Damit ergibt sich die gesuchte a posteriori-Wahrscheinlichkeit für ein Dokument d durch:

$$\begin{aligned} P(c | d) &= P(c | F) \\ &= P(F | c) \cdot P(c) \end{aligned} \quad (7.14)$$

Das letztendliche Ergebnis des Klassifikators wird dann gemäß Abschnitt 7.3.1 als Wahrscheinlichkeitsvektor W der Länge $|C|$ zurückgeliefert. Für jeden Wahrscheinlichkeitsvektor W_i eines Dokuments d_i gilt daher:

$$W_i = (P(c_1 | F_i), P(c_2 | F_i), \dots, P(c_{|C|} | F_i)) \quad (7.15)$$

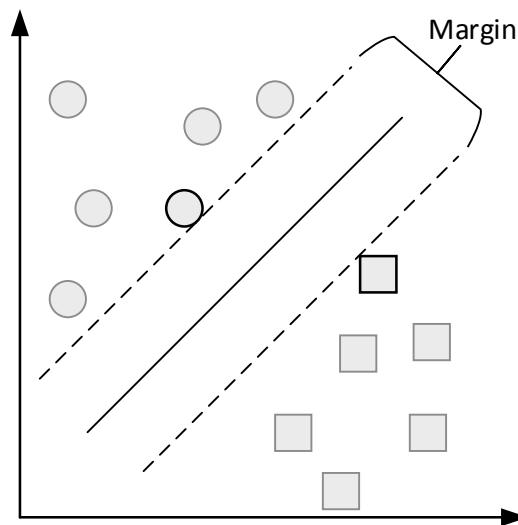


Abbildung 7.7: Vereinfachtes Beispiel zur Verdeutlichung des SVM-Klassifikators

Support Vector Machine (SVM)

Neben den Multinomial und Bernoulli Naïve Bayes-Lernalgorithmen soll in der späteren Evaluierungsphase ebenfalls ein SVM-Lernalgorithmus zum Einsatz kommen, welcher erstmals durch Vapnik und Cortes eingeführt wurde [36, 199]. Wie in Abschnitt 7.3.1 erwähnt wurde, besitzt der SVM-Klassifikator im Vergleich zum Naïve Bayes-Ansatz eine deutlich höhere Implementierungskomplexität. Eine umfassende Darstellung der Funktionsweise des SVM-Klassifikators ist im Rahmen dieser Arbeit daher nicht möglich bzw. zielführend.

Die grundlegende Idee hinter SVM soll an einem einfachen Beispiel in Abbildung 7.7 deutlich gemacht werden. Die Abbildung zeigt einen zweidimensionalen Vektorraum, welcher Instanzen aus zwei verschiedenen Klassen (als Kreise und Quadrate dargestellt) umfasst. Jeder Kreis bzw. jedes Quadrat steht für einen Feature-Vektor. In diesem einfachen Beispiel handelt es sich um linear separierbare Instanzen, d.h., die Instanzen der beiden Klassen sind durch eine Hyperebene (siehe durchgezogene Linie aus Abbildung 7.7) trennbar. Da auf dieser Ebene der größte Abstand zu den jeweiligen Klasseninstanzen herrscht, ist hier das Risiko einer fehlerhaften Klassifikation am geringsten. Das Ziel ist es daher den Abstand (engl. *margin*) zwischen den Instanzen der unterschiedlichen Klassen zu maximieren. Diejenigen Feature-Vektoren, welche am Rand der jeweiligen Klassenebenen (also an den gestrichelten Linien in Abbildung 7.7) liegen, werden als Support Vektoren bezeichnet.

SVM hat sich in den letzten Jahren trotz der hohen Implementierungskomplexität zu einem sehr populären Lernalgorithmus für Textklassifikation entwickelt [122, S. 174]. Joachims nennt verschiedene Argumente, wieso SVM-Klassifikatoren besonders gut für Textklassifikationsaufgaben geeignet sind [86, S. 139 f.]:

- Die meisten Textklassifikationsprobleme sind linear separierbar
- SVM eignet sich auch für spärlich besetzte Feature-Vektoren, d.h., Vektoren die nur wenige Elemente ungleich 0 enthalten
- SVM ist in der Lage hohe Dimensionalität bzw. umfangreiche Feature-Vektoren zu unterstützen

Die Anwendung von SVMs erfolgt typischerweise für binäre Klassifikationsprobleme, d.h. für zwei Klassen. Da es sich in der vorliegenden Problemstellung allerdings um mehr als zwei Klassen handelt, ist die Anwendung eines erweiterten Ansatzes notwendig. In der Arbeit von Hastie und Tibshirani wird diesbezüglich ein sogenannter *pairwise coupling* Ansatz empfohlen, bei dem die a posteriori Wahrscheinlichkeiten aus den binären SVM-Klassifikationsergebnissen generiert und miteinander kombiniert werden [61]. Dieser Ansatz ermöglicht die Erzeugung eines Wahrscheinlichkeitsvektors über die jeweiligen Klassen, welcher die Voraussetzung für die spätere Verwendung der Ergebnisse innerhalb des Ensemble-Klassifikators darstellt. Dieser pairwise coupling Ansatz wird im Rahmen der späteren Evaluierung verwendet.

Erweiterung der Hauptklassifikatoren

In den vorangegangenen Abschnitten wurden die Hauptlernalgorithmen sowie die Anreicherung dieser Algorithmen um die in den Rated Tags vorhandenen Bewertungskomponenten gezeigt. In diesem Abschnitt soll nun eine zusätzliche Erweiterung diskutiert werden.

Eine Herausforderung zur Gruppierung der Rated Tags ist mitunter die Kürze des Textes. Nicht selten werden für die Bewertung einzelner Aspekte nur wenige Wörter oder teilweise sogar nur ein einzelnes Wort verwendet, z.B. „komfortabel“, „Klang“ etc. Im Gegensatz zu kompletten Rezension existiert daher innerhalb der Rated Tags nur ein sehr begrenzter Kontext, der zur Ermittlung der Bedeutung eines Begriffes verwendet werden kann. Gemäß Harris bestimmt der linguistische Kontext die Ähnlichkeit von Begriffen, d.h., Begriffe, die häufiger in einem ähnlichen Kontext verwendet werden, können als ähnlich zueinander betrachtet werden [60]. Aufgrund des geringen zur Verfügung stehenden Kontextes ist vor allem die Klassifizierung von unbekannten Begriffen, d.h., Begriffen welche nicht im Vokabular der Trainingsdaten vorhanden sind ($t \notin V$), problematisch. Angenommen ein Rated Tags-Dokument d_1 mit $T_{d_1} = ('preiswert')$ wurde in der Trainingsphase mit der Klasse „Preis“ gekennzeichnet und es gilt ein neues Dokument d_2 mit $T_{d_2} = ('teuer')$ zu klassifizieren mit $t_{d_2,1} \notin V$. Obwohl die Begriffe „preiswert“ und „teuer“ eine semantische Ähnlichkeit aufweisen, ist ein traditioneller Textklassifikator nicht in der Lage Dokument d_2 korrekt zu klassifizieren, da der Begriff aus T_{d_2} nicht im Vokabular der Trainingsdaten enthalten ist.

Für die Bestimmung von semantisch ähnlichen Begriffen werden häufig Korpus- und Wissensbasierte Verfahren verwendet [134]. Ein Korpus-basiertes Verfahren ist mitunter PMI, welches bereits auf Seite 127 kurz vorgestellt wurde. Wissensbasierte Verfahren verwenden typischerweise semantische Netze, welche nach Allen und Frisch nicht nur die Informationen an sich, sondern auch das Auffinden relevanter Zusammenhänge

ermöglichen [4, S. 19]. Bekannte Vertreter dieser semantischen Netze sind beispielsweise Taxonomien oder Ontologien. Im englischsprachigen Bereich wird vor allem *WordNet* häufig als lexikalische, Domänen-neutrale Ontology eingesetzt, um Zusammenhänge zwischen Begriffen zu ermitteln [135].

Zur Adressierung des oben genannten Problems bezüglich des fehlenden Kontextes innerhalb von Rated Tags wird ein WordNet-basierter Ansatz verfolgt. Der Einsatz von WordNet im Rahmen einer Textklassifikation wurde bereits von einigen Arbeiten verfolgt. Die Ergebnisse für den Einsatz von WordNet weichen jedoch teilweise voneinander ab. Der Artikel von Jensen und Martinez zeigt, dass der Einsatz von WordNet fast immer die Klassifikationsergebnisse verbessert [85]. Auch die von Hotho et al. durchgeführten Untersuchungen haben gezeigt, dass die zusätzliche Verwendung von WordNet die Ergebnisse eines Basisklassifikators verbessert [72]. Die Arbeit von Scott und Matwin liefert gemischte Ergebnisse, welche jeweils statistisch signifikante Verbesserungen, aber auch Verschlechterungen aufzeigen [172]. Die Autoren Mansuy und Hilderman kommen zu dem Ergebnis, dass der Einsatz von WordNet nicht zu einer wesentlichen Verbesserung der Klassifikation führt [128]. Der Artikel von Hu et al. betrachtet kürzere Texte mit maximal 50 Wörtern und zeigt, dass die Hinzunahme von externen Quellen wie WordNet eine Verbesserung des Clusteringverfahrens bewirkt [78].

Die genannten Untersuchungen basieren größtenteils auf längeren Texten und sind daher nicht mit den Anforderungen von Rated Tags vergleichbar. Der Ansatz von Hu et al. [78] ist von der Länge der untersuchten Texte vergleichbar, wobei Rated Tags in der Regel wesentlich weniger als 50 Wörter enthalten. Zudem handelt es sich bei dem Ansatz von Hu et al. um einen unsupervised Ansatz, wohingegen diese Arbeit einen supervised Ansatz untersucht. Ein Beitrag dieser Arbeit ist daher die Untersuchung der Augmentierung eines Hauptklassifikators (Naïve Bayes und SVM) mithilfe von WordNet, auf Basis von sehr kurzen Bewertungstexten.

Das Vorgehen zur Augmentierung von Texten durch WordNet ist wie folgt. Vor jedem Aufruf des Hauptklassifikators wird überprüft, ob das zu klassifizierende Dokument d_i mindestens einen Begriff enthält, welcher nicht im Vokabular der Trainingsdaten enthalten ist, d.h. falls $\exists t \in T_{d_i} : t \notin V$. Für das entsprechende Dokument werden dann mittels POS-Tagging die jeweiligen Wortarten der einzelnen Wörter bestimmt, da dies eine Voraussetzung zur korrekten Interaktion mit WordNet darstellt. Für alle nicht in den Trainingsdaten enthaltenen Wörter werden dann alle Bedeutungen aus WordNet extrahiert. Zu jeder Bedeutung werden alle semantischen Verbindungen zu anderen Begriffen abgerufen. Eine Liste der semantischen Verbindungstypen ist in der Arbeit von Miller zu finden [135, S. 40], wobei Synonyme, welche in WordNet als *synonym set* (kurz *Synset*) bezeichnet werden, eine der wichtigsten Relationen darstellen.

Für alle extrahierten Begriffe aus den semantischen Relationen wird dann überprüft, ob diese ebenfalls im Vokabular V der Trainingsdaten enthalten sind. Ist dies der Fall, so kommt der Begriff als Kandidat infrage. Nach der Überprüfung aller Relationen wird der ursprüngliche, nicht im Vokabular V enthaltene Begriff durch die in der Kandidatenliste enthaltenen Begriffe ersetzt. Wurden keinerlei Kandidaten gefunden, so wird der ursprüngliche Begriff beibehalten. Der vorgeschlagene Algorithmus ist als Pseudocode in Algorithmus 7.1 zu finden.

Algorithmus 7.1: Pseudocode für das Ersetzen von nicht im Trainingsvokabular enthaltenen Wörtern mithilfe von WordNet

```

1: procedure AUGMENTDOCUMENT( $d, V$ )
2:   // Dokument mit POS-Tags versehen
3:    $taggeddocument = tagDocumentWithPOS(d)$ 
4:
5:   // Dokument in einzelne Wörter aufteilen
6:    $wordtokenlist = tokenizeToWords(taggeddocument)$ 
7:
8:   for each  $word$  in  $wordtokenlist$  do
9:     if  $word \notin V$  then
10:        $candidates = findWordNetCandidates(word, V)$ 
11:       if  $candidates.count() > 0$  then
12:          $newword = mergeToWord(candidates)$ 
13:          $word = newword$ 
14:       end if
15:     end if
16:   end for
17:   Return  $wordtokenlist$ 
18: end procedure
19:
20: procedure FINDWORDNETCANDIDATES( $wordWithPOS, V$ )
21:   // Alle Synsets aus WordNet ermitteln
22:    $synsets = findSynSets(wordWithPOS)$ 
23:
24:    $candidates = \{\}$ 
25:   for each  $synset$  in  $synsets$  do
26:     for each  $synword$  in  $synset$  do
27:       if  $synword \in V$  and  $synword \notin candidates$  then
28:          $candidates.add(synword)$ 
29:       end if
30:     end for
31:   end for
32:   Return  $candidates$ 
33: end procedure

```

Der oben genannte Algorithmus soll anhand eines Beispiels verdeutlicht werden. Angenommen ein Vokabular V enthält eine Vielzahl von Begriffen, unter anderem den Begriff „cheap“. Ein neu zu klassifizierendes Dokument d_1 enthält den Text „too expensive“. Gemäß der Vorverarbeitung der Daten (siehe Abschnitt 7.3.2) würde das Stoppwort „too“ entfernt, wodurch nur der Begriff „expensive“ erhalten bleibt. Für das Beispiel wird davon ausgegangen, dass dieser Begriff nicht im Trainingsvokabular enthalten ist. Anhand des POS-Taggings wird dann ermittelt, dass es sich bei dem Begriff „expensive“ um ein Adjektiv handelt, d.h. um das POS-Label „JJ“¹. Der mit POS-Tags versehene Text lautet daher „expensive_JJ“. Dieser wird an WordNet übergeben, um die semantischen Relationen zu ermitteln. Die Menge der potenziellen Kandidaten umfasst unter anderem $\{ 'overpriced', 'costly', 'cheap' \}$. Da der Begriff „cheap“ bereits im Trainingsvokabular vorkommt, wird der Begriff „expensive“ durch „cheap“ ersetzt. Es handelt sich in diesem Fall bei den beiden Begriffen nicht um Synonyme, sondern um Antonyme. Allerdings wird durch die semantische Verbindung davon ausgegangen, dass sie einem gemeinsamen Konzept, in diesem Beispiel dem Konzept „Preis“, angehören. Durch das Ersetzen des unbekannten Begriffes durch einen Begriff aus dem Trainingsvokabular V ist der Hauptklassifikator nun in der Lage Dokument d_1 besser zu klassifizieren.

7.3.4 String-basierte Klassifikatoren

Die zuvor genannten Naïve Bayes- und SVM-Klassifikatoren betrachten jedes Wort eines Vokabulars als eigenes Feature innerhalb des Vektors, daher werden syntaktisch ähnliche Begriffe wie beispielsweise „big“ und „bigger“ oder „picture“ und „pitcure“ als unterschiedlich angesehen. Vor allem im Rahmen von Benutzer-generierten Daten entstehen allerdings häufig Rechtschreibfehler, welche zu einer leichten Begriffsänderung führen. Hinzu kommt, dass durch die Kürze der Texte den einzelnen Wörtern eine höhere Bedeutung zukommt. Wie bereits Islam und Inkpen erwähnt haben, existieren im Vergleich zu längeren Texten weniger Forschungsarbeiten, die sich mit den Ähnlichkeiten von Kurztexten beschäftigt haben [83, S. 3]. Daher werden im folgenden Abschnitt verschiedene String-basierte Klassifikatoren vorgestellt.

Im Gegensatz zum Naïve Bayes- oder SVM-Ansatz, welche als *Eager-Learner* bezeichnet werden können, handelt es sich bei String-basierten Klassifikatoren um sogenannte *Lazy-Learner*. Der Unterschied besteht hauptsächlich darin, dass diese meistens kein explizites Modell erstellen, sondern für jede Klassifikation auf vorhandene Trainingsdaten zurückgreifen [3, S. 7]. Eager-Klassifikatoren greifen hingegen nach der Modellerstellung nicht mehr auf die Trainingsdaten, sondern nur noch auf das generierte Modell zurück.

Generelles Vorgehen

Alle nachfolgenden String-basierten Klassifikatoren verfahren nach dem gleichen Muster. Zunächst wird während der Lernphase jedes Rated Tag-Dokument, d.h. die darin vor-

¹http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html
(besucht am 06.03.2015)

kommenden Begriffe, mit den jeweils anderen Dokumenten aus den Trainingsdaten L verglichen. Für eine Dokumentenmenge L ergeben sich daher $n = \frac{|L|^2 - |L|}{2}$ Durchläufe. Für den Vergleich definiert jeder String-basierte Klassifikator k_i eine Ähnlichkeitsfunktion $\mu_{k_i}(x, y)$, wobei für die Parameter x und y während der Lernphase $x, y \in L, x \neq y$ gilt. In der späteren Klassifikationsphase wird die Ähnlichkeit zwischen nicht klassifizierten Dokumenten und den Trainingsdokumenten berechnet, daher gilt hier $x \in U, y \in L, x \neq y$. Für den Ähnlichkeitsvergleich werden jeweils die Begriffsmengen T_x und T_y der jeweiligen Dokumente miteinander verglichen. Als Ergebnis wird der Ähnlichkeitswert $\lambda_{x,y}$ mit $0 \leq \lambda_{x,y} \leq 1, \lambda_{x,y} \in \mathbb{R}$ zurückliefert. Betrachtet ein Klassifikator zwei Dokumente als komplett ähnlich zueinander, so erhält $\lambda_{x,y}$ den Wert 1.

Während der Lernphase wird zu dem ermittelten Ähnlichkeitswert $\lambda_{x,y}$ auch ermittelt, ob die Klassen c_x und c_y der verglichenen Dokumente x und y übereinstimmen. Das Ergebnis wird dann in $\varepsilon_{x,y}$ gespeichert (siehe Gleichung 7.16). Für alle $\lambda_{x,y}$ und $\varepsilon_{x,y}$ ergibt sich dann die Menge O der Länge n mit $O = \{(\lambda_{1,2}, \varepsilon_{1,2}), (\lambda_{1,3}, \varepsilon_{1,3}), \dots, (\lambda_{n-1,n}, \varepsilon_{n-1,n})\}$.

$$\varepsilon_{x,y} = \begin{cases} 1 & \text{falls } c_x = c_y \\ 0 & \text{sonst} \end{cases} \quad (7.16)$$

Nach dem Vergleich der Dokumente wird ermittelt, bei welchem Ähnlichkeitsgrenzwert τ sich der maximale F_1 -Wert¹ befindet. Hierzu wird für jeden Tupel der Menge O untersucht, ob der jeweilige Ähnlichkeitswert $\lambda_{x,y}$ zu einer korrekten Übereinstimmung der Klassen (d.h. $\varepsilon_{x,y} = 1$) geführt hat. Das Verfahren zur Ermittlung von τ wird in Algorithmus 7.2 veranschaulicht. Die in Algorithmus 7.2 (Zeile 22) verwendete Funktion *calculateF1()* basiert auf der in Abschnitt 7.4.1 dargestellten Gleichung 7.34 (siehe Seite 158). Der mittels Algorithmus 7.2 ermittelte Grenzwert τ wird dann im Rahmen der Lernphase für die spätere Klassifikation gespeichert. In der Klassifikationsphase werden zwei Dokumente x und y von einem Klassifikator k_i nur dann als ähnlich angesehen, falls für diese $\mu_{k_i}(x, y) \geq \tau$ gilt.

Weiterhin wird in der Lernphase ermittelt, welche Tupel der Menge O den Ähnlichkeitsgrenzwert τ übersteigen (siehe Gleichung 7.17). Danach wird aus O_τ die Anzahl aller Tupel ermittelt, welche zu einem korrekten Klassifikationsergebnis geführt haben. Diese Anzahl wird dann in Relation zu O_τ gesetzt, um die Gewichtung ϱ des Klassifikators zu berechnen (siehe Gleichung 7.18). Die Gewichtung steht für die Wahrscheinlichkeit, dass ein über dem Grenzwert τ liegender Ähnlichkeitswert zu einer korrekten Klassifikation führt, wobei $0 \leq \varrho \leq 1, \varrho \in \mathbb{R}$.

$$O_\tau = \{o \mid o \in O \wedge o_\lambda \geq \tau\} \quad (7.17)$$

$$\varrho = \begin{cases} \frac{|\{o \in O_\tau \mid o_\varepsilon = 1\}|}{|O_\tau|} & \text{falls } |O_\tau| > 0 \\ 0 & \text{sonst} \end{cases} \quad (7.18)$$

¹Siehe Abschnitt 7.4.1 für die Erläuterung des F_1 -Wertes und der weiteren Evaluierungsmetriken

Algorithmus 7.2: Pseudocode für die Ermittlung des Ähnlichkeitswertes bei dem der F_1 -Wert sein Maximum erreicht

```

1: procedure GETSIMILARITYWITHMAXF1( $O$ )
2:    $\tau = 0.0$  // Ähnlichkeitswert bei dem der  $F_1$ -Wert maximal ist
3:    $maxF1 = 0.0$ 
4:
5:   for each  $similarity$  in  $O_\lambda$  do
6:      $truePositive = 0$  // TP
7:      $falsePositive = 0$  // FP
8:      $falseNegative = 0$  // FN
9:
10:    for each  $(\lambda, \varepsilon)$  in  $O$  do
11:      if  $\varepsilon == 1$  then
12:        if  $\lambda \geq similarity$  then
13:           $truePositive = truePositive + 1$ 
14:        else
15:           $falseNegative = falseNegative + 1$ 
16:        end if
17:      else if  $\lambda \geq similarity$  then
18:         $falsePositive = falsePositive + 1$ 
19:      end if
20:    end for
21:
22:     $F1 = calculateF1()$  // basierend auf TP, FP und FN
23:
24:    if  $F1 > maxF1$  then
25:       $maxF1 = F1$ 
26:       $\tau = similarity$ 
27:    end if
28:  end for
29:
30:  Return  $\tau$ 
31: end procedure

```

Da die Ergebnisse aller Klassifikatoren später durch einen Ensemble-Klassifikator aggregiert werden sollen (siehe Abschnitt 7.3.5), liefert jeder String-basierte Klassifikator für jedes zu klassifizierende Dokument nicht die wahrscheinlichste Klasse, sondern einen Wahrscheinlichkeitsvektor der Länge $|C|$ zurück¹. In diesem werden die Ähnlichkeitswerte für die einzelnen Klassen aufsummiert. Es wird ermittelt, ob die Ähnlichkeit $\lambda_{x,y}$ zwischen einem nicht klassifizierten Dokument x und einem Dokument der Trainingsdaten y den Ähnlichkeitsgrenzwert τ übersteigt. Ist dies der Fall, so wird im Wahrscheinlichkeitsvektor der Wert für c_y um $\lambda_{x,y}$ inkrementiert. Nach dem Vergleich mit allen Trainingsdokumenten wird der Wahrscheinlichkeitsvektor normalisiert, so dass die wahrscheinlichste Klasse den Wert 1 erhält. Abschließend werden die Elemente des normalisierten Wahrscheinlichkeitsvektors mit der in der Lernphase berechneten Gewichtung ϱ multipliziert, wodurch sich der finale Wahrscheinlichkeitsvektor ergibt.

Ermittlung der Ähnlichkeit von Texten

Die wohl am nächsten liegende und einfachste Methode für den Vergleich zweier Texte ist die Überprüfung auf Gleichheit. Für diesen Zweck wird für den Klassifikator k_{Equals} eine Ähnlichkeitsfunktion $\mu_{k_{Equals}}(x, y)$ definiert (siehe Gleichung 7.19). Die Ähnlichkeitsfunktion vergleicht die Wortmengen zweier Dokumente, d.h., die Wortreihenfolge wird im Rahmen der Ähnlichkeitsbetrachtung nicht berücksichtigt. Besteht Gleichheit zwischen den beiden Wortmengen T_x und T_y so wird der Wert 1, ansonsten der Wert 0 zurückgeliefert.

$$\mu_{k_{Equals}}(x, y) = \begin{cases} 1 & \text{falls } T_x = T_y \\ 0 & \text{sonst} \end{cases} \quad (7.19)$$

Für die Analyse von nicht komplett identischen Texten gibt es verschiedene Verfahren zur Bestimmung der Ähnlichkeit. Eine Möglichkeit ist die Überprüfung, ob ein Teil eines Textes innerhalb eines anderen Textes vorkommt, z.B. das Wort „weight“ innerhalb des Wortes „lightweight“. Basierend auf der Länge der beiden Texte lässt sich dann ein normalisierter Wert ermitteln, der die Ähnlichkeit ausdrückt. Eine weitere Metrik ist die längste gemeinsame Teilfolge zwischen zwei Texten, auch *Longest Common Subsequence (LCS)* genannt [5, S. 305]. Hier müssen allerdings die in den Texten vorkommenden Teilfolgen nicht konsekutiv sein. So wäre beispielsweise die längste gemeinsame Teilfolge für die Wörter „peace“ und „price“ die Zeichenfolge „pce“.

In den Bereich der Teilfolgen fällt ebenfalls die sogenannte *N-Gramm-Analyse*. Hier werden die zu vergleichenden Wörter in aufeinanderfolgende Fragmente der Länge N aufgeteilt. Zur Verdeutlichung sollen die oben genannten Wörter „peace“ und „price“ anhand der Methode *NGram₂* in N-Gramme der Länge 2 (auch *Bigramme* genannt) zerlegt werden:

$$\begin{aligned} NGram_2('peace') &= \{'pe', 'ea', 'ac', 'ce'\} \\ NGram_2('price') &= \{'pr', 'ri', 'ic', 'ce'\} \end{aligned} \quad (7.20)$$

¹Vgl. Gleichung 7.3 auf Seite 132

Basierend auf den sich ergebenden Mengen aus Gleichung 7.20 können dann verschiedene Ähnlichkeiten berechnet werden, z.B. der *Jaccard*- oder *Dice-Koeffizient* (siehe Gleichung 7.21).

$$\begin{aligned} Jaccard_n(s_1, s_2) &= \frac{|NGram_n(s_1) \cap NGram_n(s_2)|}{|NGram_n(s_1) \cup NGram_n(s_2)|} \\ Dice_n(s_1, s_2) &= \frac{2 \cdot |NGram_n(s_1) \cap NGram_n(s_2)|}{|NGram_n(s_1)| + |NGram_n(s_2)|} \end{aligned} \quad (7.21)$$

Ein weiteres Mittel zur Bestimmung von Ähnlichkeiten sind *Distanzfunktionen*. Je kleiner die Distanz zwischen zwei Begriffen s_1 und s_2 ist, desto ähnlicher sind sich diese. Einer der wichtigsten Distanzmetriken ist die *Editierdistanz* (engl. *edit distance*), bei welcher die Distanz anhand der Anzahl an Editieroperationen (z.B. das Einfügen und Löschen von Zeichen), welche zur Überführung von s_1 nach s_2 notwendig sind, berechnet wird [34, S. 74]. Wagner und Fischer bezeichnen dies auch als *String-to-String correction Problem* [204, S. 168]. Wird das Ergebnis der Editierdistanzfunktion $\delta(s_1, s_2)$ auf ein Intervall $[0, 1]$ normalisiert, so kann daraus die Ähnlichkeit der Begriffe s_1 und s_2 wie folgt abgeleitet werden: $\mu(s_1, s_2) = 1 - \delta(s_1, s_2)$. Eine Editierdistanz ist vor allem im Hinblick auf die Erkennung von Rechtschreibfehlern und damit besonders für Benutzer-generierte Inhalte interessant.

Nach der Betrachtung von bestehenden Algorithmen zur Ermittlung der Ähnlichkeit zweier Texte werden nachfolgend die innerhalb des Klassifikationsmodells verwendeten bzw. neu entwickelten Algorithmen vorgestellt. Der erste vorgeschlagene Klassifikator umfasst die Analyse von Teilfolgen innerhalb von Wörtern. Hierzu wurde die von Islam und Inkpen beschriebene Methodik für die Erstellung einer Ähnlichkeitsmatrix adaptiert [83]. Anstelle der von Islam und Inkpen vorgeschlagenen Ähnlichkeitsalgorithmen wird jedoch ein eigener Algorithmus verwendet, welcher nachfolgend erläutert wird.

Für die Darstellung wird davon ausgegangen, dass zwei zu vergleichende Wortmengen T_x und T_y existieren, wobei $|T_x| \leq |T_y|$. Die Wortmengen T_x und T_y werden dann in einzelne Wörter zerlegt. Danach wird eine Matrix M_{T_x, T_y} der Größe $|T_x| \times |T_y|$ erstellt, welche später für jedes Wortpaar die Ähnlichkeit $\lambda_{t_{x_i}, t_{y_j}}$ enthält. Diese Ähnlichkeit wird berechnet, indem jedes Wort t_{x_i} mit jedem Wort t_{y_j} verglichen wird. Bei dem Vergleich wird ermittelt, ob das kürzere der beiden Wörter in dem längeren Wort enthalten ist. Für einen normalisierten Wert zwischen 0 und 1 wird die Länge der Überschneidung dann durch die Länge des längeren Wortes dividiert. Eine formale Darstellung hierzu ist in Gleichung 7.22 zu finden.

$$\lambda_{t_{x_i}, t_{y_j}} = \begin{cases} \frac{|t_{x_i} \cap t_{y_j}|}{|t_{y_j}|} & \text{falls } |t_{x_i}| \leq |t_{y_j}|, t_{x_i} \in t_{y_j} \\ \frac{|t_{x_i} \cap t_{y_j}|}{|t_{x_i}|} & \text{falls } |t_{x_i}| > |t_{y_j}|, t_{y_j} \in t_{x_i} \\ 0 & \text{sonst} \end{cases} \quad (7.22)$$

Die ermittelte Ähnlichkeit wird dann in die Matrix M_{T_x, T_y} übernommen. Nach dem Vergleich aller Wörter der beiden Mengen, wird die gesamte Ähnlichkeit durch Aggregation der einzelnen Ähnlichkeiten ermittelt. Das Vorgehen hierzu ist wie folgt. Aus der Matrix M_{T_x, T_y} wird das Element $m_{t_{x_i}, t_{y_j}}$ mit der maximalen Ähnlichkeit ermittelt. Die gesamte Ähnlichkeit $\lambda_{x,y}$ wird um diesen Betrag erhöht und die Zeile i sowie die Spalte j in der sich das Element befindet werden aus der Matrix entfernt. Dieser Vorgang wird so lang wiederholt, bis die Anzahl der Matrixzeilen den Wert 0 erreicht hat. Für eine Normalisierung wird der ermittelte Ähnlichkeitswert $\lambda_{x,y}$ durch die ursprüngliche Breite der Matrix $|T_y|$ dividiert.

Der oben beschriebene Vorgang soll an einem kurzen Beispiel verdeutlicht werden. Angenommen es existieren zwei Kurztexte „ease of setup“ und „easy to set up“. Nach der Entfernung von Stoppwörtern (hier „of“ und „to“) würden sich folgende Wortmengen T_x und T_y ergeben:

$$\begin{aligned} T_x &= \{ 'ease', 'setup' \} \\ T_y &= \{ 'easy', 'set', 'up' \} \end{aligned} \quad (7.23)$$

Ein Vergleich der beiden Mengen ergibt die Matrix aus Gleichung 7.24, wobei nur die Wortpaare („setup“, „set“) und („setup“, „up“) einen Ähnlichkeitswert größer als 0 ergeben haben. Dadurch ergibt sich eine Ähnlichkeit von $\frac{3}{5}$, welche im Rahmen der Normalisierung in Relation zur Länge $|T_y|$ gesetzt wird, wodurch sich die endgültige Ähnlichkeit von $\frac{1}{5}$ ergibt (vgl. Gleichung 7.25).

$$M_{T_x, T_y} = \begin{matrix} & \begin{matrix} easy & set & up \end{matrix} \\ \begin{matrix} ease \\ setup \end{matrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{2}{5} \end{pmatrix} \end{matrix} \quad (7.24)$$

$$\lambda_{x,y} = \frac{\frac{3}{5} + 0}{3} = \frac{1}{5} \quad (7.25)$$

Eine Schwäche des obigen Ansatzes ist, dass eine Ähnlichkeit nur erkannt wird, falls ein Wort komplett in einem anderen Wort enthalten ist. Die obigen Wörter „ease“ und „easy“ werden daher nicht als ähnlich angesehen, obwohl die ersten drei Zeichen der Wörter identisch sind. Um diesen Sachverhalt einzubeziehen, wird ein weiterer Algorithmus vorgeschlagen, der die längste gemeinsame aufeinanderfolgende Teilfolge beginnend bei dem ersten Zeichen zurückliefert (siehe Algorithmus 7.3). Zwei Wörter werden bei diesem Ansatz bewusst nur dann als ähnlich erkannt, falls sie mit der gleichen Zeichenfolge beginnen. Dadurch sollen falsch positiv Ergebnisse reduziert werden, d.h., der Ansatz fokussiert sich auf die Verbesserung der Genauigkeit und nicht auf die Trefferquote. Für die Berechnung der gesamten Ähnlichkeit zweier Wortmengen wird ebenfalls das oben dargestellte Matrix-basierte Verfahren von Islam und Inkpen angewandt.

Als letzter String-basierter Algorithmus soll der Einsatz der *Levenshtein-Distanz*, einer bekannten Editierdistanz, gezeigt werden. Wie bereits erwähnt wurde, eignen sich Editierdistanzen besonders zur Erkennung von Rechtschreibfehlern. Verwendet

Algorithmus 7.3: Pseudocode für die Ermittlung der Ähnlichkeit zweier Wörter basierend auf identischer Anfangszeichenfolge

```

1: procedure COMMONSTARTSIMILARITY( $s1, s2$ )
2:    $commoncharacters = 0$ 
3:
4:   for  $i = 1$  to  $Min(|s1|, |s2|)$  do
5:     //  $i$  steht für das  $i$ -te Zeichen des jeweiligen Wortes
6:     if  $s1_i == s2_i$  then
7:        $commoncharacters = commoncharacters + 1$ 
8:     else
9:       Exit For
10:    end if
11:  end for
12:
13:   $similarity = commoncharacters / Max(|s1|, |s2|)$ 
14:  Return  $similarity$ 
15: end procedure

```

ein Benutzer beispielsweise das Wort „pucture“ anstelle von „picture“, so hätte die Editierdistanz den Wert 1, da genau eine Austauschoperation ($u \mapsto i$) notwendig ist, um „pucture“ nach „picture“ zu überführen. Im Rahmen des Klassifikationsmodells wird eine sehr konservative Umsetzung des Levenshtein-Ansatzes verfolgt, d.h., der Fokus liegt in der Optimierung der Genauigkeit (engl. *precision*) und nicht der Trefferquote (engl. *recall*)¹. Daher werden zwei Wörter nur als ähnlich betrachtet, falls deren Editierdistanz maximal den Wert 1 beträgt, d.h., wenn für die Transformation nur maximal eine Editieroperation notwendig ist. Als Editieroperationen werden das Einfügen, Löschen, Ersetzen und Vertauschen von Zeichen betrachtet. Das Vertauschen von zwei Zeichen, welches häufig durch Tippfehler zustande kommt, wird als eine Editieroperation gezählt. Ein Beispiel wäre das falsch geschriebene Wort „piculture“. Durch das Vertauschen der Zeichen „t“ und „c“ kann es in die korrekte Form „picture“ überführt werden.

Für die Berechnung der Ähnlichkeit wird die Levenshtein-Distanz zunächst in Relation zu der Länge des Kürzeren der beiden Wörter gesetzt. Durch die Subtraktion dieses Wertes vom Wert 1 ergibt sich die Ähnlichkeit der beiden Wörter (siehe Gleichung 7.26). Entsprechend den obigen Algorithmen wird dieser Vergleich für alle Wortpaare zweier Wortmengen durchgeführt und deren Ähnlichkeit anhand der Ähnlichkeitsmatrix bestimmt.

$$\lambda_{t_{x_i}, t_{y_j}} = \begin{cases} 1 - \frac{\delta(t_{x_i}, t_{y_j})}{Min(|t_{x_i}|, |t_{y_j}|)} & \text{falls } Min(|t_{x_i}|, |t_{y_j}|) > 0, \delta(t_{x_i}, t_{y_j}) \leq 1 \\ 0 & \text{sonst} \end{cases} \quad (7.26)$$

¹Vgl. Abschnitt 7.4.1 für die Metriken Precision und Recall

7.3.5 Ensemble-Klassifikator

Nach der Darstellung der in den vorigen Abschnitten genannten Basisklassifikatoren soll in diesem Abschnitt ein Ensemble-Klassifikator vorgestellt werden, welcher die Ergebnisse der Basisklassifikatoren kombiniert und daraus das finale Klassifikationsergebnis ermittelt. Wie in diversen Arbeiten postuliert wird, liefern Ensemble-Klassifikatoren bessere Ergebnisse als einzelne Basisklassifikatoren alleine [99, 143, 157].

Für die Kombination der Klassifikationsergebnisse ist die Art und Weise relevant, in der die Basisklassifikatoren ihr Ergebnis zurückliefern. Gemäß Xu et al. lassen sich die Ausgaben von verschiedenen Klassifikationsalgorithmen in die folgenden drei Ebenen unterteilen [215, S. 420]:

1. **Abstrakte Ebene:** Ein Klassifikator liefert lediglich die vorhergesagte Klasse zurück.
2. **Rang-basierte Ebene:** Ein Klassifikator dieser Ebene liefert eine Rangfolge aller Klassen zurück, wobei das Element mit Rang 1 der ersten Wahl des Klassifikators entspricht.
3. **Messbare Ebene:** Ein Klassifikator liefert für jede Klasse einen numerischen Wert zurück, welcher die Zuversicht hinsichtlich der Klassifikation widerspiegelt.

Wie bereits mehrfach erwähnt wurde, wird für die oben genannten Basisklassifikatoren die messbare Ebene verwendet, bei der jede Klassifikationsfunktion einen Wahrscheinlichkeitsvektor zurückliefert, der die Wahrscheinlichkeiten für jede Klasse enthält. Dies ist unter anderem in Tabelle 7.4 dargestellt. Die Aufgabe eines Ensemble-Klassifikators $\gamma_E(u)$ umfasst die Analyse der einzelnen Wahrscheinlichkeitsvektoren und die nachfolgende Ermittlung der wahrscheinlichsten Klasse $c \in C$ für ein nicht klassifiziertes Dokument $u \in U$, d.h. $\gamma_E : u \mapsto c$. Für diese Arbeit wird also ein Ensemble-Klassifikator gewählt, welcher die Ergebnisse der Basisklassifikatoren kombiniert und somit als nicht-trainierbarer Ensemble-Klassifikator kategorisiert werden kann [107, 157].

Zunächst stellt sich die Frage wie die einzelnen Wahrscheinlichkeitsvektoren kombiniert werden. Da es sich bei den Wahrscheinlichkeiten um reelle Zahlen handelt, kommen diverse algebraische Kombinationsmethoden infrage. Zwei typische Regeln zur Kombination sind die Durchschnittsregel sowie die Maximierungsregel. Bei ersterer Variante wird aus den Wahrscheinlichkeiten aller Klassifikatoren eine Durchschnittswahrscheinlichkeit gebildet. Bei der Maximierungsregel wird die höchste Wahrscheinlichkeit zurückgeliefert. Für diese Arbeit wird für die Kombination der Wahrscheinlichkeiten die Durchschnittsregel verwendet, da jeder Klassifikator gleichmäßig gewichtet und somit die Dominanz eines Klassifikators reduziert werden soll. Die diesbezügliche Aggregationsfunktion $\eta(c_i, u)$ wird in Gleichung 7.27 veranschaulicht. Sie ermittelt die durchschnittliche Wahrscheinlichkeit einer Klasse c_i für ein nicht klassifiziertes Dokument u , basierend auf den Wahrscheinlichkeitsvektoren aller Klassifikatoren K .

$$\eta(c_i, u) = \frac{1}{|K|} \cdot \sum_{n=1}^{|K|} w_{k_n c_i u} \quad (7.27)$$

Tabelle 7.4: Wahrscheinlichkeitsvektoren der Basisklassifikatoren und Aggregationsvektor $\Omega(u)$ des Ensemble-Klassifikators

$m = K , n = C $ $u \in U$	c_1	c_2	...	c_n
$\gamma_{k_1}(u)$	$w_{k_1 c_1 u}$	$w_{k_1 c_2 u}$...	$w_{k_1 c_n u}$
$\gamma_{k_2}(u)$	$w_{k_2 c_1 u}$	$w_{k_2 c_2 u}$...	$w_{k_2 c_n u}$
\vdots	\vdots	\vdots	\vdots	\vdots
$\gamma_{k_m}(u)$	$w_{k_m c_1 u}$	$w_{k_m c_2 u}$...	$w_{k_m c_n u}$
$\Omega(u)$	$\eta(c_1, u)$	$\eta(c_2, u)$...	$\eta(c_n, u)$

Die Anwendung der obigen Aggregationsfunktion $\eta(c_i, u)$ auf alle Klassen liefert dann den Wahrscheinlichkeitsvektor $\Omega(u)$ (vgl. Gleichung 7.28 sowie Tabelle 7.4). Da die Wahrscheinlichkeiten aller Basisklassifikatoren auf das Intervall $[0, 1]$ normalisiert wurden, ergibt sich auch für die Wahrscheinlichkeiten des Vektors $\Omega(u)$ jeweils ein Wert zwischen 0 und 1.

$$\Omega(u) = (\eta(c_1, u), \eta(c_2, u), \dots, \eta(c_{|C|}, u)) \quad (7.28)$$

Um die Klasse mit der höchsten Wahrscheinlichkeit zu ermitteln, werden zunächst diejenigen Klassen ermittelt, welche in $\Omega(u)$ den maximalen Wert aufweisen. Diese Menge X_u (siehe Gleichung 7.29) kann unter Umständen mehrere Elemente enthalten, falls verschiedene Klassen das gleiche Maximum besitzen. Die endgültige Klassifikationsfunktion $\gamma_E(u)$ des Ensemble-Klassifikators liefert dann ungeachtet davon, immer nur das erste Element dieser Menge (x_{u_1} aus Gleichung 7.29) zurück.

$$\begin{aligned} X_u &= \{c \mid c \in C \wedge \eta(c, u) = \arg \max \Omega(u)\} \\ \gamma_E(u) &= x_{u_1} \end{aligned} \quad (7.29)$$

7.4 Evaluierung

Nachdem im vorigen Abschnitt die Einzelheiten des Klassifikationsmodells erläutert wurden, soll nun in diesem Abschnitt dessen Evaluierung aufgezeigt werden. Zu Beginn wird in Abschnitt 7.4.1 das methodische Vorgehen der Untersuchung beschrieben, welches unter anderem die Beschaffung der Daten sowie die der Evaluierung zugrunde liegenden Metriken verdeutlicht. Anschließend werden die Ergebnisse der Untersuchung in Abschnitt 7.4.2 präsentiert und in Abschnitt 7.4.3 diskutiert sowie mit den Ergebnissen verwandter Arbeiten verglichen.

7.4.1 Methodisches Vorgehen

Datenbeschaffung

Für eine aussagekräftige Evaluierung sind echte Benutzer-generierte Tags in hoher Anzahl erforderlich. Aufgrund der Neuartigkeit von Rated Tags liegen solche Daten jedoch noch nicht vor. Da in der Evaluierung primär die Vereinheitlichung von Tags und weniger die Bewertungskomponente eine Rolle spielt, werden daher bestehende Daten des Formats F_{PK} verwendet (vgl. Seite 125). Das heißt, es werden Benutzer-generierte Pro- und Kontrapunkte verwendet. Die Bewertungskomponente umfasst daher zwei mögliche Ausprägungen, nämlich 0 (Kontra) und 1 (Pro).

Für die Beschaffung der Daten kamen verschiedene Möglichkeiten infrage. Eine erste Überlegung war die Verwendung eines bestehenden Datensatzes¹ aus der Arbeit von Liu et al. [119]. Leider konnte dieser jedoch für die Evaluierung nicht verwendet werden, da die darin enthaltenen Pro- und Kontra-Datensätze nicht nach Produktkategorien getrennt vorliegen und somit für ein Klassifikationsverfahren ungeeignet sind. Um eine möglichst heterogene Auswahl an Produktkategorien zu evaluieren, wurde eine Bewertungsseite gesucht, welche Benutzer-generierte Bewertungen für eine breite Produktpalette bereitstellt. Ein solches Bewertungsportal ist Buzzillions, welches laut eigenen Angaben mehr als 17 Millionen Rezensionen umfasst².

Für die Evaluierung wurden aus dem Bewertungsportal Buzzillions die folgenden fünf Produktkategorien ausgewählt:

- Digitalkamera
- Global Positioning System (GPS)-Gerät
- Fernseher
- Laufschuhe
- Kindersitz (Auto)

Der Schwerpunkt der Kategorien ist aus dem Bereich der Elektroartikel, da hier ein größeres Potenzial zu erwarten ist als bei Produktkategorien, welche typischerweise weniger Attribute aufweisen. Auf der Webseite von Buzzillions wurden die Produkte jeder Kategorie zunächst absteigend nach der Anzahl der Benutzerrezensionen sortiert. Aus den obersten Einträgen wurden dann jeweils zufällig sechs Produkte pro Kategorie ausgewählt. Für die Extraktion der jeweiligen Pro- und Kontra-Tags aus den Rezensionen wurde ein Webcrawler entwickelt, der automatisiert alle Rezensionen nach dem Vorkommen solcher Tags untersucht. Eine beispielhafte Darstellung solcher Pro- und Kontra-Tags der Webseite Buzzillions ist Abbildung 7.8 zu entnehmen. Wie dort zu erkennen ist, werden die Tags jeweils durch ein Komma separiert.

Nach der Extraktion der Rezensionen samt Tags hat sich gezeigt, dass für die Kategorien „Schuhe“ und „Kindersitz“ im Vergleich zu den anderen Kategorien deutlich weniger

¹<http://www.cs.uic.edu/~liub/FBS/pros-cons.rar> (besucht am 17.02.2015)

²<http://www.buzzillions.com/about> (besucht am 16.03.2015)

PROS: Easy to setup, Good connections, Good soundtrack support, Great picture quality

CONS: Slow load time



[Add Comment](#)

[Report Abuse](#)

Abbildung 7.8: Beispielhafte Darstellung der vergebenen Pro- und Kontra-Tags eines Benutzers

Tabelle 7.5: Übersicht der extrahierten Daten

Kategorie	Tags (unterschiedlich)	Tags (insgesamt)	Rezensionen
Kamera	939	5203	924
GPS	1419	15016	2543
Fernseher	1480	7861	1849
Schuhe	947	20473	4382
Kindersitz	441	15778	2714
<i>Summe</i>	<i>5226</i>	<i>64331</i>	<i>12412</i>

verschiedenartige Tags vorhanden waren. Daher wurden für diese beiden Kategorien nochmals jeweils sechs Produkte nach dem oben genannten Verfahren ausgewählt. Die Rezensionen und Tags dieser Produkte wurden anschließend extrahiert und mit den bestehenden Daten kombiniert. Als Datenbasis für die spätere Verarbeitung wurde eine relationale Datenbank (Microsoft SQL Server 2014 Express) verwendet.

Die endgültige Anzahl an Tags aller Kategorien ist Tabelle 7.5 zu entnehmen. Für die spätere Evaluierung der Vereinheitlichung der Daten sind die unterschiedlichen Tags relevant. Die Spalte „Tags (insgesamt)“ zeigt die Häufigkeit der Verwendung aller Tags. Hieraus lässt sich ableiten, dass Konsumenten dazu neigen Tags wiederzuverwenden, wobei der Grad der Wiederverwendung in den Kategorien teilweise stark voneinander abweicht.

Erstellung eines Gold Standards

Damit in der späteren Evaluierungsphase bestimmt werden kann, ob die vorgestellten Klassifikatoren korrekt arbeiten, muss jeder Tag vorab manuell klassifiziert werden, d.h., jedem Tag wird eine eindeutige Klasse zugewiesen. Dieser Prozess wird auch als Erstellung eines sogenannten *Gold Standards*, oder im englischen auch als *Ground Truth*, bezeichnet [127, S. 152].

Hierzu ist für jede Produktkategorie zunächst die Definition von relevanten Klassen notwendig. Es wurde eine möglichst feine Aufteilung der Klassen vorgenommen, wodurch

Tabelle 7.6: Anzahl der unterschiedlichen Klassen je Produktkategorie

Kategorie	Anzahl Klassen
Kamera	23
GPS	21
Fernseher	23
Schuhe	15
Kindersitz	13

eine hohe Klassenanzahl resultiert. Beispielsweise wurden die Aspekte „Größe“ und „Gewicht“ als separate Klassen betrachtet und nicht zu einem Aspekt „Größe und Gewicht“ gruppiert. Es ist zu beachten, dass die Erstellung solcher Klassen immer subjektiver Natur ist und unterschiedliche Betreiber von Rated Tags unterschiedliche Anforderungen hinsichtlich des Detailgrades der Klassen besitzen. Tabelle 7.6 zeigt die jeweilige Anzahl der Klassen für die Produktkategorien. Die maximale Anzahl von Klassen einer Kategorie liegt bei 23. Dies mag zwar sehr hoch erscheinen, ist aber nicht als unüblich anzusehen. In der Studie von Zhai et al. wurden für die Kategorie „Staubsauger“ sogar 28 Klassen definiert [219, S. 1277].

Zusätzlich zu den in Tabelle 7.6 gezeigten Klassen wurde für jede Kategorie eine Klasse erzeugt, welche für nicht relevante Tags verwendet wird. Es ist davon auszugehen, dass Anwender auch Tags vergeben, welche keine Produktattribute adressieren, z.B. die Tags „Love this TV“ oder „Great camera“. Solche Bewertungen beziehen sich auf das Produkt selbst und nicht auf ein Produktmerkmal. Diese allgemeine Bewertung wird bereits durch die Gesamtbewertung des Rezensenten ausgedrückt. Die vorgestellten Klassifikationsverfahren müssen also auch in der Lage sein, solche nicht relevanten Inhalte korrekt als solche zu klassifizieren.

Nach der Erzeugung der Klassen für die Produktkategorien wurde jeder Tag aus den extrahierten Daten mit einer Klasse versehen. Insgesamt wurden für die Untersuchung 5226 Tags klassifiziert (vgl. Tabelle 7.5). Auf diese Weise kann später überprüft werden, ob die vom Klassifikator ermittelte Klasse eines Tags mit seiner tatsächlichen Klasse übereinstimmt.

Evaluierungsmetriken

Um die Güte des Klassifikators zu messen, haben sich in der Vergangenheit verschiedene Evaluierungsmetriken bewährt. Im Fall einer binären Klassifikation (wie beispielsweise bei einem Spamfilter) gehört ein Dokument entweder zu der positiven oder der negativen Klasse ($C = \{ 'positiv', 'negativ' \}$). Im vorliegenden Fall existieren mehrere Klassen, d.h., es gibt eine positive Klasse und n negative Klassen mit $n = |C| - 1$. Diese n negativen Klassen können zu einer negativen Klasse zusammengefasst werden, wodurch wieder ein binäres Klassifikationsproblem entsteht. In der nachfolgenden Darstellung

Tabelle 7.7: Wahrheitsmatrix zur Verdeutlichung der Evaluierungsmetriken

		Tatsächlicher Wert	
		Positiv	Negativ
Berechneter Wert	Positiv	TP	FP
	Negativ	FN	TN

beschreibt die Notation c eine Klasse der Menge C , wohingegen \bar{c} ein Element des Komplements aus C und c darstellt, d.h. $\bar{c} \in (C \setminus c)$. Durch die Überführung in ein binäres Klassifikationsproblem ergibt sich mit $C' = \{c, \bar{c}\}$ die Menge der möglichen Klassen. Basierend darauf sind für einen Klassifikationsprozess die folgenden vier Ergebnisse möglich:

- **True positive (TP):**
Ein Dokument der Klasse c wird als c klassifiziert.
- **False positive (FP):**
Ein Dokument der Klasse \bar{c} wird als c klassifiziert.
- **True negative (TN):**
Ein Dokument der Klasse \bar{c} wird als \bar{c} klassifiziert.
- **False negative (FN):**
Ein Dokument der Klasse c wird als \bar{c} klassifiziert.

Zur weiteren Verdeutlichung der oben genannten Metriken ist in Tabelle 7.7 eine 2×2 -Kontingenztafel bzw. Wahrheitsmatrix (engl. *confusion matrix*) dargestellt, welche die vier Ausprägungen nochmals verdeutlicht.

Die vier oben genannten Metriken bilden die Basis für weitere Evaluierungsmetriken. Eine bekannte Metrik ist die *Präzision* (engl. *accuracy*) eines Klassifikators (siehe Gleichung 7.30). Allerdings spiegelt die Präzision nicht immer die Robustheit eines Modells wieder. Dies gilt vor allem für Daten für die eine nicht gleichmäßige Klassenverteilung vorliegt [118, S. 81]. Dies wird am Beispiel eines Spamfilters deutlich. Angenommen 90 % aller zu klassifizierenden E-Mails sind nicht Spam, dann würde ein Klassifikator, der keinerlei Berechnungen vornimmt und immer die Klasse „Nicht-Spam“ zurückliefert, eine Präzision von etwa 90 % aufweisen.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7.30)$$

Um die genannte Problematik der Präzisionsmetrik zu vermeiden, bieten sich die Evaluierungsmetriken Genauigkeit (engl. *precision*) und Trefferquote (engl. *recall*) an. Die Genauigkeit berechnet sich, indem die korrekt klassifizierten positiven Instanzen (TP) in Relation zu allen als positiv klassifizierten Instanzen (TP und FP) gesetzt werden (siehe Gleichung 7.31). Für die Trefferquote wird anstelle von FP, also dem Fehler 1. Art, die Metrik FN, also der Fehler 2. Art verwendet (siehe Gleichung 7.32).

$$P = \frac{TP}{TP + FP} \quad (7.31)$$

$$R = \frac{TP}{TP + FN} \quad (7.32)$$

Obwohl die Metriken Genauigkeit und Trefferquote nicht direkt in Relation zueinander stehen, ist es in der Praxis häufig so, dass eine Erhöhung der einen Metrik zur Reduzierung der anderen Metrik führt [118, S. 82]. Daher wird die sogenannte *F-Wert*-Metrik verwendet, welche auf den Metriken Genauigkeit und Trefferquote basiert. Die Berechnung des F-Wertes ist Gleichung 7.33 zu entnehmen, wobei β für den Gewichtungsfaktor steht. Ein typischer Wert für β ist 1, wodurch sich der häufig verwendete F_1 -Wert ergibt. Hier werden Genauigkeit und Trefferquote gleich gewichtet, d.h., es wird das harmonische Mittel beider Werte gebildet (siehe Gleichung 7.34).

$$F_\beta = \left(1 + \beta^2\right) \cdot \frac{P \cdot R}{\beta^2 \cdot P + R} \quad (7.33)$$

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (7.34)$$

Bei dem vorliegenden Klassifikationsproblem, welches mehrere Klassen adressiert, bestehen zwei mögliche Varianten zur Berechnung der obigen Metriken: Microaveraging und Macroaveraging. Microaveraging aggregiert auf Dokumentenebene alle einzelnen Klassifikationsergebnisse in einer eigenen Wahrheitstabelle und berechnet daraus die o.g. Metriken. Beim Macroaveraging hingegen wird zunächst für jede Klasse die jeweilige Metrik bestimmt und anschließend der Mittelwert all dieser Werte gebildet. Gemäß Manning et al. werden beim Microaveraging häufig vorkommende Klassen höher gewichtet, während beim Macroaveraging weniger vorkommende Klassen stärker berücksichtigt werden [127, S. 281]. Gemäß Wiener et al. ist der Microaverage F_1 -Wert aufgrund der Gewichtung von häufiger vorkommenden Klassen als optimistisch einzustufen [212, S. 326]. Da im vorliegenden Klassifikationsproblem unausgeglichene Klassengrößen erwartet werden, weil populäre Produktaspekte häufiger genannt werden als sekundäre Merkmale, werden im Rahmen der Evaluierung die Macroaverage Werte der genannten Metriken berichtet.

Für die Ermittlung der oben genannten Metriken wird eine 10-fache stratifizierte Kreuzvalidierung (engl. *10-fold stratified cross validation*) durchgeführt. Gemäß Cimiano wird eine Kreuzvalidierung typischerweise zur Ermittlung der statistischen Robustheit eines Modells verwendet [32, S. 64]. Bei einer k -fachen stratifizierten Kreuzvalidierung werden per Zufallsprinzip alle vorhandenen Dokumente D in k (hier $k = 10$) gleich große Dokumentenmengen unterteilt. Durch die Stratifizierung wird darauf geachtet, dass die Klassen innerhalb aller k Dokumentenmengen annähernd so verteilt sind wie es in der gesamten Dokumentenmenge der Fall ist. Der Kreuzvalidierungsprozess wird insgesamt k mal durchgeführt, wobei immer $k - 1$ Dokumentenmengen als Trainingsdaten und

die verbleibende Dokumentenmenge als Testdaten verwenden werden. Dies bedeutet, dass nach der Durchführung der Kreuzvalidierung, jede der k Dokumentenmengen genau einmal getestet wurde. Anhand der Ergebnisse der einzelnen Durchläufe werden dann die Mittelwerte gebildet, um die Güte des Klassifikators zu bestimmen. Der Prozess der 10-fachen stratifizierten Kreuzvalidierung wird insgesamt fünfmal wiederholt, wobei für jeden Durchlauf eine andere zufällige Aufteilung der k Dokumentenmengen vorgenommen wird. Über die Ergebnisse der fünf Durchläufe wird dann ein Mittelwert gebildet.

Für die Durchführung der Kreuzvalidierung wurde die Java-basierte Data Mining Software WEKA¹ in der Version 3.6.12 verwendet. Die verwendeten Hauptklassifikatoren Bernoulli Naïve Bayes², Multinomial Naïve Bayes³ und SVM⁴ entstammen ebenfalls dem WEKA-Framework. Die in Abschnitt 7.3 entworfenen Lernalgorithmen bzw. Klassifikatoren wurden für den Einsatz in WEKA auf Basis der Programmiersprache Java entwickelt.

7.4.2 Ergebnisse

Datenbasis der Evaluierung

Bevor die Klassifikationsergebnisse der Evaluierung dargestellt werden, sollen zunächst die Charakteristiken der verwendeten Daten aufgezeigt werden. Eine Übersicht der extrahierten Rezensionen sowie der dazugehörigen Tags ist in Abbildung 7.9 dargestellt. Wie dort zu erkennen ist, weisen die Kategorien „Schuhe“ und „Kindersitz“ trotz der doppelten Anzahl an ausgewählten Produkten im Durchschnitt immer noch weniger verschiedenartige Tags auf als die Produkte aus dem Elektronikbereich. Ein kausaler Zusammenhang ist jedoch nicht nachweisbar. Es könnte daran liegen, dass für diese beiden Produktgruppen weniger relevante verschiedenartige Aspekte vorhanden sind als beispielsweise bei den anderen drei Produktgruppen aus dem Elektronikbereich. Eine andere Möglichkeit wäre, dass Konsumenten der Produktkategorien „Schuhe“ und „Kindersitz“ eher dazu geneigt sind bestehende Tags wiederzuverwenden, anstatt neue Tags zu erstellen.

Hinsichtlich der Wiederverwendung von Tags zeigt sich, dass Konsumenten durchaus geneigt sind bestehende Tags zu verwenden. Tabelle 7.8 zeigt die zehn häufigsten Tags für jede Produktkategorie. Aus der Tabelle ist ersichtlich, dass es einen Konsens zwischen Konsumenten bezüglich der relevanten Tags gibt. Ferner enthält keiner der in Tabelle 7.8 dargestellten Tags einen Rechtschreibfehler. Dies unterstützt die These, dass Konsumenten auch ohne eine kontrollierte Instanz in der Lage sind, sich auf wichtige Produkthaspekte zu einigen.

¹<http://www.cs.waikato.ac.nz/ml/weka/> (besucht am 16.03.2015)

²<http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayes.html> (besucht am 16.03.2015)

³<http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayesMultinomial.html> (besucht am 16.03.2015)

⁴<http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SMO.html> (besucht am 16.03.2015)

Tabelle 7.8: Übersicht der zehn häufigsten Tags pro Kategorie

(a) Fernseher		(b) GPS	
Tag	Anzahl	Tag	Anzahl
Great picture quality	1031	Easy to read	1680
Attractive	582	Compact	1675
Good remote control	533	Easy to set up	1636
Easy to setup	528	Easy menus	1482
Good audio	506	Simple controls	1402
Energy efficient	438	Acquires satellites quickly	1380
Great resolution	388	Reliable	1356
Reduces glare	388	Large clear lcd	956
Easy to set up	340	Acquires satellites slowly	331
Good connections	318	Short battery life	244

(c) Kamera		(d) Kindersitz	
Tag	Anzahl	Tag	Anzahl
Compact	538	Durable	2150
Easy to use	511	Comfortable	2126
Good in low light	438	Safe to use	2002
Large clear lcd	388	Easy to install	1885
Great resolution	386	Adjustable	1754
Fast shutter	241	Stylish	1547
Great picture quality	229	Easy to clean	1130
Short battery life	222	Heavy / bulky	827
Zoom range/performance	153	Lightweight	667
Image stabilization	140	Difficult to adjust	251

(e) Schuhe	
Tag	Anzahl
Comfortable	3813
Good cushioning	2934
Durable	2174
Lightweight	2165
Arch support	2121
Attractive	1208
Pronation control	1086
Stylish	1037
Great traction	859
Breathable	591

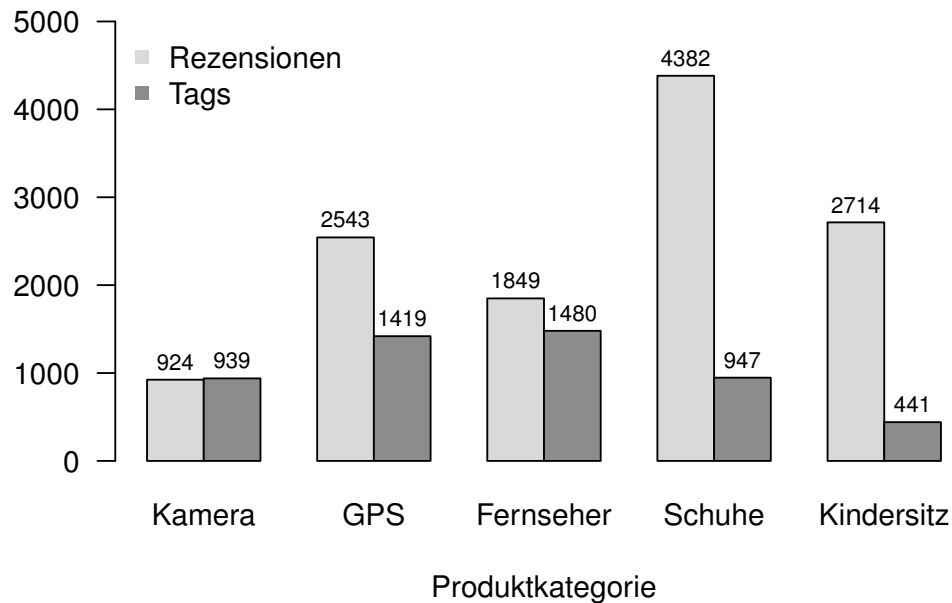


Abbildung 7.9: Anzahl der extrahierten Rezensionen sowie der Anteil der darin enthaltenen verschiedenartigen Tags

Die Analyse der vorhandenen Tags zeigt allerdings auch, dass ein sogenannter *Long-Tail* entsteht, bei dem viele unterschiedliche Tags und nur eine begrenzte Anzahl an stabilen Tags existieren [59, S. 220]. Diese Verteilung wird in Abbildung 7.10 verdeutlicht. Für jede der fünf Produktkategorien existieren etwa 20 Tags, welche regelmäßig von Anwendern verwendet werden. Der Großteil der vergebenen Tags wird allerdings nur einmal verwendet. Aufgrund der Menge an unterschiedlichen Tags wurde die Anzahl in Abbildung 7.10 auf 100 beschränkt, um die Lesbarkeit der Grafik zu gewährleisten.

Die Ergebnisse aus Abbildung 7.10 sind vergleichbar mit denen von Dellschaft und Staab. Auch dort wurde der Einsatz von Tags untersucht, allerdings für den Social Bookmarking Dienst Delicious¹. Die Studie von Dellschaft und Staab zeigt, dass die Häufigkeit nach dem siebten bis zehnten Tag stark abnimmt [40, S. 79]. Die Hypothese von Dellschaft und Staab ist, dass dieses Verhalten darauf beruht, dass Anwendern in Delicious nur die sieben häufigsten Tags angezeigt werden. Aus Abbildung 7.10 ist ersichtlich, dass im Rahmen der eigenen Studie ein Abfall der Häufigkeiten durchschnittlich nach dem zehnten Tag stattfindet. Die Benutzeroberfläche von Buzzillions schlägt dem Anwender jeweils die acht häufigsten Pro- und Kontra-Tags vor. Ein kausaler Zusammenhang zwischen dem Abfall und der Anzahl der vorgeschlagenen Tags ist allerdings auch hier nicht ermittelbar.

Einige Arbeiten im Bereich des Social Tagging haben die oben genannten Häufigkeitsverteilungen untersucht. Allerdings basierten diese Studien immer auf klassischen Tagging-

¹<http://www.delicious.com> (besucht am 15.03.2015)

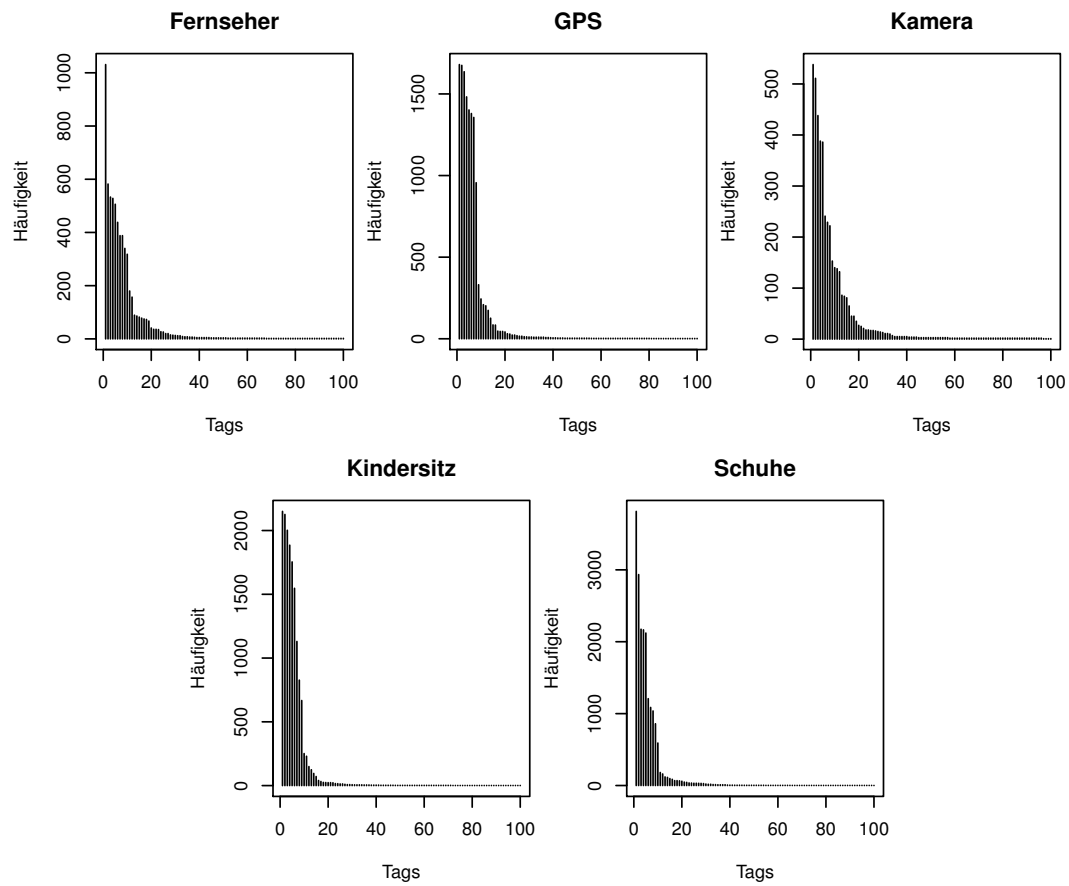


Abbildung 7.10: Erkennung eines Long-Tails bei der Häufigkeitsverteilung der Tags für alle Produktkategorien

Tabelle 7.9: Übersicht der Mediane für die Anzahl der Wörter und Zeichen der verwendeten Tags

	Ohne Vorverarbeitung	Mit Vorverarbeitung
Wörter eines Tags (Median)	3	2
Zeichen eines Tags (Median)	20	11

systemen (wie beispielsweise Delicious), welche primär zur Inhaltsklassifikation genutzt werden. Ein Forschungsbeitrag der vorliegenden Arbeit ist die Untersuchung eines Taggingsystems, bei dem Benutzermeinungen in Form von Tags verwendet werden. Nach aktuellem Kenntnisstand existiert keine weitere Untersuchung die sich mit diesem Thema beschäftigt hat. Als Ergebnis lässt sich festhalten, dass auch Benutzermeinungen in Form von Tags den Potenzgesetzen aus traditionellen Taggingssystemen folgen und Anwender einen Konsens hinsichtlich der verwendeten Tags erzielen.

Ein weiteres Merkmal der untersuchten Tags ist deren Datenknappheit. Wie Tabelle 7.9 zeigt, beträgt der Median für die Anzahl der Wörter pro Tag nur drei. Nach der Vorverarbeitung der Daten (z.B. aufgrund der Entfernung von Stoppwörtern) beträgt der Median sogar nur noch zwei Wörter pro Tag. Ähnliches gilt für die Anzahl der Zeichen eines Tags. Hier beträgt der Median nach der Vorverarbeitung nur noch elf Zeichen. Dies verdeutlicht das Problem der Datenknappheit im Hinblick auf die Klassifikation.

Da viele der in Abschnitt 7.2.2 betrachteten Ansätze, z.B. [25, 56, 73], für Produktmerkmale nur Substantive heranziehen, ist die Verteilung der Wortarten innerhalb der Tags interessant. Hierzu wurden die POS-Informationen nach der Vorverarbeitung der Daten ausgewertet. Es zeigt sich, dass Tags in 83,98 % der Fälle Substantive enthalten. Adjektive sind in 43,13 % der Tags zu finden. Verben und Adverbien werden nur in 25,03 % bzw. 7,12 % der Fälle verwendet. Anhand dieser Zahlen zeigt sich die Schwäche der oben genannten Ansätze. Da diese nur Substantive für die Erkennung von Produktmerkmalen verwenden, wird ein großer Teil der Tags von Anfang an nicht erfasst.

Klassifikationsergebnisse

Nach der allgemeinen Betrachtung der zugrunde liegenden Daten sollen nun die Klassifikationsergebnisse des entwickelten Ensemble-Klassifikators untersucht werden. Als Ausgangswert für spätere Vergleiche dient zum einen das Klassifikationsergebnis ohne Vorverarbeitung. Zum anderen wird das Klassifikationsergebnis unter Verwendung einer traditionellen Vorverarbeitung verglichen, d.h., es werden alle Schritte aus Abschnitt 7.3.2 mit Ausnahme der folgenden, speziell für das Modell entwickelten Ansätze durchgeführt:

- Keine Normalisierung von Währungsbeträgen und Zeiteinheiten
- Keine Entfernung von neutralen Adjektiven
- Keine Entfernung von negativen Präfixen

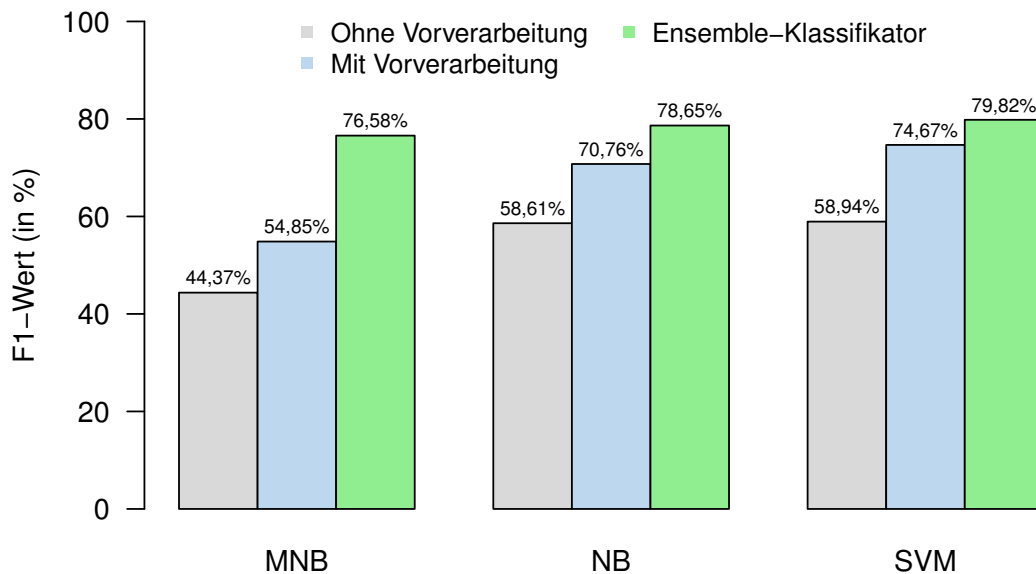


Abbildung 7.11: Durchschnittliche F_1 -Werte der Hauptklassifikatoren mit und ohne Verwendung des Ensemble-Klassifikators (MNB=Multinomial Naïve Bayes, NB=Bernoulli Naïve Bayes, SVM=Support Vector Machine)

Die durchschnittlichen F_1 -Werte der 10-fachen stratifizierten Kreuzvalidierung sind in Abbildung 7.11 dargestellt. Die detaillierten Ergebnisse, aufgeteilt nach Produktkategorien, sind den Tabellen 7.10 und 7.11 zu entnehmen. Dort werden jeweils die gewichtete Genauigkeit (Precision), die gewichtete Trefferquote (Recall) sowie der Macroaverage F_1 -Wert (jeweils in Prozent) berichtet. In Tabelle 7.10a ist die Leistung der drei Hauptklassifikatoren Bernoulli Naïve Bayes, Multinomial Naïve Bayes und SVM bei der Klassifikation der Daten ohne Vorverarbeitung dargestellt. Tabelle 7.10b zeigt die Ergebnisse der Hauptklassifikatoren unter Verwendung einer traditionellen Vorverarbeitung der Daten wie im vorigen Absatz beschrieben. Die Tabelle 7.10b zeigt deutlich, dass der SVM-Klassifikator gegenüber den Naïve Bayes-basierten Ansätzen eine höhere Klassifikationsgüte aufweist.

Die Klassifikationsgüte des in dieser Arbeit entwickelten Ensemble-Klassifikators ist den Tabellen 7.11a und 7.11b zu entnehmen. Auch hier zeigt sich, dass der SVM-Klassifikator der am besten geeignete Hauptklassifikator ist, wobei die Unterschiede zu den Naïve Bayes-basierten Klassifikatoren geringer ausfallen als in Tabelle 7.10b. Um den konkreten Mehrwert des Ensemble-Klassifikators hervorzuheben, enthält Tabelle 7.11b einen Vergleich zwischen den Klassifikationsergebnissen mit und ohne Ensemble-Klassifikator basierend auf den F_1 -Werten des SVM-Hauptklassifikators. Im Durchschnitt übersteigt der F_1 -Wert des Ensemble-Klassifikators den alleinigen Einsatz von SVM um 5,15 %. Ausgehend von dem F_1 -Wert von 74,67 % beträgt die relative Steigerung des F_1 -Wertes durch den Ensemble-Klassifikator daher 6,9 %.

Im Hinblick auf die Naïve Bayes-basierten Klassifikatoren ergibt sich durch den Einsatz des Ensemble-Klassifikators noch eine deutlichere Steigerung der Klassifikationsgüte. Der

Tabelle 7.10: Evaluierungsergebnisse der Hauptklassifikatoren (MNB=Multinomial Naïve Bayes, NB=Bernoulli Naïve Bayes, SVM=Support Vector Machine)

(a) Klassifikationsergebnisse ohne jegliche Vorverarbeitung

Kategorie	Precision			Recall			F ₁ -Wert		
	MNB	NB	SVM	MNB	NB	SVM	MNB	NB	SVM
Fernseher	56,82 %	63,16 %	68,82 %	53,27 %	64,31 %	64,05 %	46,01 %	62,13 %	65,20 %
Kamera	56,19 %	59,14 %	66,27 %	47,88 %	56,34 %	58,87 %	43,49 %	55,66 %	61,09 %
GPS	64,56 %	61,34 %	68,68 %	51,31 %	60,87 %	53,05 %	49,13 %	60,60 %	56,74 %
Kindersitz	52,99 %	61,80 %	61,76 %	47,30 %	58,14 %	58,23 %	42,29 %	57,33 %	58,81 %
Schuhe	57,30 %	58,59 %	60,12 %	47,73 %	58,37 %	50,56 %	40,94 %	57,30 %	52,83 %
Durchschnitt	57,57 %	60,81 %	65,13 %	49,50 %	59,61 %	56,95 %	44,37 %	58,61 %	58,94 %

(b) Klassifikationsergebnisse nach traditioneller Vorverarbeitung der Daten

Kategorie	Precision			Recall			F ₁ -Wert		
	MNB	NB	SVM	MNB	NB	SVM	MNB	NB	SVM
Fernseher	62,32 %	72,46 %	77,78 %	58,78 %	70,88 %	76,09 %	52,66 %	70,19 %	76,38 %
Kamera	64,42 %	72,83 %	77,53 %	58,42 %	71,10 %	75,78 %	55,35 %	70,47 %	75,73 %
GPS	73,07 %	72,53 %	75,47 %	62,44 %	72,30 %	73,05 %	60,50 %	71,93 %	73,24 %
Kindersitz	65,19 %	70,58 %	76,92 %	56,92 %	67,89 %	72,24 %	54,48 %	68,07 %	73,05 %
Schuhe	68,03 %	75,09 %	77,25 %	55,88 %	73,69 %	75,18 %	51,25 %	73,13 %	74,97 %
Durchschnitt	66,61 %	72,70 %	76,99 %	58,49 %	71,17 %	74,47 %	54,85 %	70,76 %	74,67 %

Tabelle 7.11: Evaluierungsergebnisse des entworfenen Ensemble-Klassifikators

(a) Ergebnisse der Hauptklassifikatoren unter Verwendung des Ensemble-Klassifikators

Kategorie	Precision			Recall			F ₁ -Wert		
	MNB	NB	SVM	MNB	NB	SVM	MNB	NB	SVM
Fernseher	83,55 %	84,70 %	85,76 %	83,76 %	85,20 %	85,84 %	82,88 %	84,59 %	85,63 %
Kamera	78,34 %	79,90 %	80,90 %	77,74 %	79,91 %	80,49 %	76,45 %	78,86 %	80,02 %
GPS	75,03 %	76,46 %	77,77 %	74,60 %	76,70 %	77,02 %	74,07 %	76,25 %	76,83 %
Kindersitz	77,34 %	75,94 %	80,22 %	75,83 %	75,51 %	79,32 %	75,59 %	75,19 %	79,23 %
Schuhe	76,74 %	78,94 %	78,24 %	74,72 %	78,97 %	77,82 %	73,90 %	78,34 %	77,39 %
Durchschnitt	78,20 %	79,19 %	80,58 %	77,33 %	79,26 %	80,10 %	76,58 %	78,65 %	79,82 %

(b) Darstellung des Mehrwerts des Ensemble-Klassifikators basierend auf SVM

Kategorie	F ₁ -Werte bei Verwendung des SVM-Klassifikators		
	Ohne Vorverarbeitung	Mit Vorverarbeitung	Ensemble-Klassifikator
Fernseher	65,20 %	76,38 %	85,63 %
Kamera	61,09 %	75,73 %	80,02 %
GPS	56,74 %	73,24 %	76,83 %
Kindersitz	58,81 %	73,05 %	79,23 %
Schuhe	52,83 %	74,97 %	77,39 %
Durchschnitt	58,94 %	74,67 %	79,82 %

F_1 -Wert bei Einsatz von Bernoulli Naïve Bayes steigt dadurch von 70,76 % auf 78,65 %, was einer relativen Steigerung von 11,15 % entspricht. Beim Einsatz des Multinomial Naïve Bayes findet sogar eine noch drastischere Verbesserung der Güte statt. Der F_1 -Wert erhöht sich hier von 54,85 % auf 76,58 %, und entspricht somit einer relativen Steigerung von 39,62 %.

7.4.3 Diskussion

Bewertung der Ergebnisse

Die Klassifikationsergebnisse der Evaluierung sind vielversprechend. Durch den Einsatz des entwickelten Ensemble-Klassifikators konnte der durchschnittliche F_1 -Wert der Hauptklassifikatoren teilweise deutlich gesteigert werden. Die Ergebnisse bezüglich der Hauptklassifikatoren entsprechen den genannten Vermutungen. Der SVM-Klassifikator hat für die vorliegende Problemstellung die beste Klassifikationsgüte aller Hauptklassifikatoren erreicht. Der Bernoulli Naïve Bayes ist dem Multinomial Naïve Bayes ohne Verwendung des Ensemble-Klassifikators deutlich überlegen (vgl. Tabelle 7.10b). Dies war zu vermuten, da die Stärke des Bernoulli Naïve Bayes bei der Klassifikation von kurzen Texten bereits bekannt ist [127, S. 268]. Jedoch hat sich gezeigt, dass durch den Einsatz des Ensemble-Klassifikators der Multinomial Naïve Bayes eine annähernd ähnliche Klassifikationsgüte erreichen kann wie der Bernoulli Naïve Bayes. Generell ist festzuhalten, dass durch den Einsatz des Ensemble-Klassifikators nur noch ein relativ geringer Unterschied zwischen der Verwendung von Naïve Bayes-basierten Ansätzen und SVM existiert.

Die Untersuchung hat im Hinblick auf die heterogenen Produktkategorien gezeigt, dass der Ensemble-Klassifikator für alle Kategorien annähernd gleich gute Ergebnisse geliefert hat. Die beste Güte wurde für die Kategorie „Fernseher“ erreicht. Dies liegt jedoch vor allem daran, dass für diese Kategorie die meisten Tags zur Verfügung standen (vgl. Abbildung 7.9). Die Kategorie „GPS“ hat überraschenderweise den geringsten F_1 -Wert erzielt, obwohl hier die zweitmeisten Trainingsdaten zur Verfügung standen. Ein Grund hierfür könnte sein, dass sich die Funktionsweise des GPS-Gerätes teilweise schwer anhand von Schlagwörtern beschreiben lässt. Aus den verwendeten Daten ist ersichtlich, dass teilweise sehr verschiedene Formulierungen verwendet wurden, um die gleichen Aspekte des GPS-Gerätes zu beschreiben, z.B. „Crazy routes“, „Sometimes just wrong“, „Says street after passing“ oder „Not enough reminders“. Diese Beispiele machen deutlich, dass sich die Heterogenität der Formulierungen negativ auf die Leistung des Klassifikators auswirkt.

Vergleich mit verwandten Arbeiten

Wie bereits im Rahmen der Abgrenzung diskutiert wurde, existieren nicht viele Arbeiten die sich mit der Gruppierung von Produktaspekten basierend auf Kurztexten bzw. Tags beschäftigen haben. Aus den in Tabelle 7.1 auf Seite 129 dargestellten Ansätzen ist die Arbeit von Liu et al. [119] direkt mit dieser Arbeit vergleichbar. Auch dort wurden

Produktraspekte basierend auf kurzen Benutzermeinungen extrahiert und gruppiert. Allerdings wurde in der Arbeit von Liu et al. ein sehr konservativer Ansatz zur Gruppierung verwendet. Der Ansatz erzielte eine Genauigkeit von 100 % bei einer Trefferquote von 52 % [119, S. 350]. Dies resultiert in einem F_1 -Wert von 68,42 %. Der durchschnittliche F_1 -Wert des in dieser Arbeit vorgestellten Ensemble-Klassifikators, unter Verwendung von SVM, beträgt 79,82 %. Ein 1:1-Vergleich beider Arbeiten ist jedoch leider nicht möglich, da die der Evaluierung zugrunde liegenden Daten unterschiedlich sind. Leider ist aus dem Artikel von Liu et al. auch nicht ersichtlich wie viele Trainingsdaten verwendet wurden („We manually tagged a large collection of reviews“ [119, S. 349]). Im Gegensatz zu dieser Arbeit verwendeten Liu et al. eine eher homogene Datenbasis, welche nur Elektronikprodukte umfasst. Ferner geht aus deren Arbeit leider nicht hervor, welche konkreten Produktkategorien verwendet wurden.

Ein weiterer Beitrag der vorliegenden Arbeit bezieht sich auf die String-basierte Klassifikation von kurzen Texten. Wie sich gezeigt hat, führt die Ergänzung der Hauptklassifikatoren um String-basierte Klassifikatoren im Rahmen des Ensemble-Klassifikators zu einer Verbesserung der Klassifikationsgüte. Zum Vergleich der String-basierten Klassifikatoren wurden die von Islam und Inkpen vorgeschlagenen Algorithmen implementiert und anstelle der in Abschnitt 7.3.4 entworfenen String-basierten Klassifikatoren verwendet. Die Arbeit von Islam und Inkpen konzentriert sich ebenfalls auf Kurztexte, wobei deren Evaluierung kurze Sätze und Phrasen umfasst. Eine Schwäche des Ansatzes von Islam und Inkpen ist, dass für jede verwendete Ähnlichkeitsfunktion der gleiche Gewichtungsfaktor verwendet wird, wodurch die Stärken und Schwächen der Ähnlichkeitsfunktionen nicht angemessen in den Algorithmus mit einfließen. Dagegen werden in dieser Arbeit die Gewichtungen der einzelnen Ähnlichkeitsfunktionen in der Lernphase anhand des maximalen F_1 -Wertes ermittelt.

Für den Vergleich der String-basierten Algorithmen dieser Arbeit mit denen aus Islam und Inkpen wurde der entwickelte Ensemble-Klassifikator (unter Verwendung von SVM) als Basis genommen. Die resultierenden F_1 -Werte beider Ansätze sind in Tabelle 7.12 dargestellt. Die Ergebnisse zeigen, dass die entwickelten String-basierten Ansätze für alle Produktkategorien etwas bessere F_1 -Werte liefern als die Algorithmen von Islam und Inkpen. Es ergibt sich eine relative Verbesserung des F_1 -Wertes von 2,14 %. Da sich die Klassifikation von sehr kurzen Texten aufgrund der Informationsknappheit sehr schwierig gestaltet, kann auch diese geringfügige Verbesserung der F_1 -Werte als positiv angesehen werden.

Ausblick

Im Rahmen der durchgeführten Untersuchung wurden insgesamt fünf heterogene Produktkategorien betrachtet. Die Ergebnisse hierzu sind vielversprechend. Im Hinblick auf eine Generalisierung der Ergebnisse wäre jedoch eine zukünftige Evaluierung von weiteren Produktkategorien interessant.

Es ist zu beachten, dass die durchgeführte Evaluierung auf englischen Tagging-Daten basiert. In einer zukünftigen Studie könnte untersucht werden, inwiefern sich die Ergebnisse für Tags einer anderen Sprache, z.B. Deutsch, unterscheiden. Der Großteil der für

Tabelle 7.12: Vergleich der String-basierten Klassifikatoren dieser Arbeit mit der Arbeit von Islam und Inkpen [83]

Kategorie	F₁-Werte (SVM-Klassifikator)	
	Islam und Inkpen [83]	Diese Arbeit
Fernseher	83,21 %	85,72 %
Kamera	78,32 %	79,95 %
GPS	74,95 %	76,84 %
Kindersitz	77,56 %	79,23 %
Schuhe	76,72 %	77,38 %
Durchschnitt	78,15 %	79,82 %

den Ensemble-Klassifikator entwickelten Algorithmen ist sprachneutral gestaltet, z.B. die String-basierten Klassifikatoren. Anstelle der in dieser Arbeit verwendeten lexikalischen Datenbank WordNet könnte in einer Studie zur Klassifikation von deutschsprachigen Tags das deutsche Pendant GermaNet¹ genutzt werden.

Im Rahmen einer zukünftigen Untersuchung könnte ferner evaluiert werden, ob eine zusätzliche Textanalyse der mit den Tags verbundenen Rezensionen das Klassifikationsergebnis verbessert. Durch den Einbezug der Rezensionstexte könnte zusätzlicher Kontext gewonnen werden, der bei der Klassifikation der wenigen Taginformationen behilflich sein könnte. Für die Analyse der Rezensionstexten (Review Mining) existieren bereits einige Ansätze (vgl. unter anderem Tabelle 7.1 auf Seite 129), welche hierfür verwendet werden könnten.

¹<http://www.sfs.uni-tuebingen.de/GermaNet> (besucht am 20.03.2015)

Kapitel 8

Service-orientierte Bereitstellung von Rated Tags

In den vorangegangenen Kapiteln 4 bis 6 wurde zunächst das Konzept von Rated Tags vorgestellt, prototypisch implementiert und anhand einer Anwenderstudie evaluiert. Ferner wurde im vorhergehenden Kapitel ein Klassifikationsmodell zur Vereinheitlichung der entstehenden Tags entwickelt, implementiert und anhand einer realen Datenbasis evaluiert. In diesem Kapitel soll nun die Service-orientierte Bereitstellung des Rated Tags-Ansatzes betrachtet werden. Diese Bereitstellungsart mit der Bezeichnung „Rated Tags as a Service“ soll bestehenden Onlineshops oder Bewertungsportalen eine einfache Möglichkeit bieten, die Funktionalitäten des Rated Tags-Ansatzes zu nutzen, ohne die notwendige Infrastruktur dafür bereitzustellen bzw. zu verwalten. Nachfolgend werden zunächst die mit der Service-orientierten Bereitstellung verbundenen Ziele diskutiert. Für die Service-orientierte Bereitstellung werden vor allem die Ansätze aus Sicht des Dienstgebers betrachtet, wie etwa die Datenhaltung oder die Schnittstellenkommunikation, sowie eine Integration des Dienstes aus Sicht der Dienstnehmer. Ferner werden Möglichkeiten zur Dienstnehmer-übergreifenden Nutzung der erstellten Tags diskutiert, um Synergieeffekte zu erreichen und Kaltstartprobleme zu reduzieren.

8.1 Zielsetzung und Motivation

Die grundlegende Idee von *Rated Tags as a Service (RTaaS)* ist die Service-orientierte Bereitstellung des Rated Tags-Ansatzes für KMU aus dem Bereich E-Commerce, z.B. Onlinehändler oder Bewertungsportale. Diesen Anbietern soll eine einfache Integration des Rated Tags-Ansatzes ermöglicht werden, um deren Anwendern wiederum eine detaillierte Bewertungsmöglichkeit sowie eine Entscheidungshilfe zur Verfügung zu stellen. Dieses Zusammenspiel wird in Abbildung 8.1 verdeutlicht. Die Abbildung zeigt die jeweiligen Endanwender der Entscheidungshilfe. Diese Konsumenten interagieren mit Onlinehändlern bzw. Bewertungsportalen, welche als Dienstnehmer klassifiziert werden können. Diese wiederum greifen auf die Schnittstellen eines Dienstgebers zu, um die Funktionalität von Rated Tags zu integrieren.

Ein Ziel des Service-orientierten Bereitstellungskonzepts für Rated Tags ist die damit gesteigerte Wiederverwendbarkeit des Ansatzes im Vergleich zu einer komponentenorientierten Bereitstellung. Würde das Rated Tags-Modell in Komponentenform (z.B. als Softwareframework) bereitgestellt werden, so würde dies einen festen Bezug zu einer

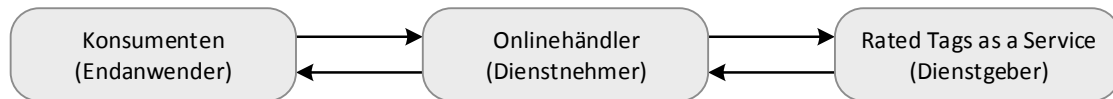


Abbildung 8.1: Konzeptuelles Modell der Dienst-orientierten Bereitstellung des Rated Tags-Ansatzes

konkreten Programmiersprache bzw. Plattform darstellen. Dies würde die Wiederverwendbarkeit reduzieren, da eine Integration in existierende Softwarelösungen nur möglich wäre, wenn die Kompatibilität der Plattformen gewährleistet ist. Im Vergleich dazu soll über eine Dienst-orientierte Bereitstellung eine Technologie-unabhängige Einbindung ermöglicht werden.

Ein weiterer Vorteil der dienstorientierten Bereitstellung ist ein geringerer Administrationsaufwand seitens der Dienstnehmer. Für die Nutzung des Dienstes muss keine neue Hardware (z.B. Festplatten etc.) oder Software (z.B. Lizenzkosten für Betriebssysteme oder Datenbanksysteme) beschafft werden. Zudem fallen für die Wartung der Software sowie deren Aktualisierung keine zusätzlichen Kosten an. Durch die Auslagerung der Funktionalität müssen sich Dienstnehmer also weniger mit der benötigten Infrastruktur befassen, sondern können sich mehr auf ihr Kerngeschäft konzentrieren. Dieser Aspekt ist vor allem für KMUs interessant.

8.2 Einordnung

Die Bereitstellung von Funktionalitäten in Dienst-orientierter Weise kann dem Themenkomplex des *Service-Oriented Computing (SOC)* zugeordnet werden. Papazoglou et al. beschreiben SOC wie folgt:

„Service-Oriented Computing (SOC) is a new computing paradigm that utilizes services as the basic constructs to support the development of rapid, low-cost and easy composition of distributed applications even in heterogeneous environments.“ [149, S. 223]

Aus oben genannter Definition ist für dieses Kapitel besonders der Aspekt der Komposition von heterogenen Anwendungen relevant, da der Service-orientierte Entwurf des Rated Tags-Ansatzes für Servicenehmer verschiedenster Plattformen anwendbar sein soll. Der Aufbau solcher Dienst-orientierter Systeme wird typischerweise dem Themenbereich *SOA* zugeordnet. SOA kann auf einer abstrakten Ebene als Architekturstil bezeichnet werden, bei dem das System aus Dienstnehmern und Dienstgebern besteht [15, S. 3]. Den zentralen Aspekt stellt ein Dienst dar.

Ein weiterer relevanter Themenbereich im Rahmen einer Service-orientierten Bereitstellung von Software ist das Konzept *SaaS*. Gemäß Buxmann et al. handelt es sich bei SaaS um eine Internet-basierte Standardsoftware, welche Kunden als Dienstleistung bereitgestellt wird [24, S. 500]. Die Nutzung einer SaaS-Lösung findet typischerweise über einen Webbrowser statt und ist damit Client-seitig nicht an eine konkrete Plattform

gebunden [24, 169]. Ein bekanntes Beispiel für eine solche SaaS-Lösung ist die Webseite [Salesforce.com](http://www.salesforce.com)¹, welche eine Customer-Relationship-Management (CRM)-Lösung gegen ein entsprechendes Nutzungsentgelt zur Verfügung stellt.

Anstelle eine Software zu kaufen und innerhalb der eigenen Infrastruktur zu installieren und zu verwalten, wird also im Rahmen von SaaS die Software als Dienstleistung bezogen. Buxmann et al. sprechen bei SaaS daher auch von einem Mietmodell, da Anwender für die gemieteten Serviceleistungen Nutzungsgebühren bezahlen [24, S. 500 f.]. Dieses Modell ist ebenfalls für RTaaS anwendbar.

8.3 Dienstarchitektur

In diesem Abschnitt sollen die beteiligten Subsysteme des Service-orientierten Modells sowie deren Entwurfsaspekte genauer betrachtet werden. Auf oberster Abstraktionsebene lassen sich drei Akteure bzw. Systeme identifizieren. Diese wurden bereits in Abbildung 8.1 im Rahmen des konzeptuellen Modells zusammengefasst. Es handelt sich dabei um Konsumenten der Onlinehändler, d.h. um die Endanwender der Entscheidungshilfe. Diese Konsumenten interagieren daher mit den Systemen der Onlinehändler und haben keine Kenntnisse davon, ob bzw. welche weiteren Systeme zur Bereitstellung der Entscheidungshilfe herangezogen werden. Die Onlinehändler wiederum greifen auf den RTaaS-Dienst zurück und nehmen daher die Rolle des Dienstnehmers ein. Die verbleibende Entität ist der Dienstgeber, welcher den Dienst bereitstellt und verwaltet. Auf hoher Ebene kann die Architektur daher als Client-Server Architektur bezeichnet werden, bei der die Systeme der Onlinehändler in dieser Konstellation sowohl die Rolle des Clients als auch des Servers einnehmen.

8.3.1 Bereitstellungsmodell

Wie bereits in Abschnitt 8.2 erwähnt wurde, besitzt RTaaS einige Charakteristiken von SaaS, bezieht sich aber im Gegensatz dazu nicht auf die Bereitstellung einer kompletten Anwendung, sondern auf die Bereitstellung von Funktionalitäten zur Integration mit bestehenden Inhalten. Genau wie bei SaaS stellt sich bei RTaaS die Frage nach der Art des Bereitstellungsmodells. Aufgrund der Kosteneffizienz und der damit verbundenen Attraktivität für KMU besteht im Rahmen einer SaaS-basierten Anwendung häufig der Bedarf nach einer Mandantenfähigkeit [55, S. 551]. Der Begriff *Mandantenfähigkeit* (engl. *multi-tenancy*) steht für einen Ansatz zur Bereitstellung von Software, bei dem sich verschiedene Mandanten eine gemeinsame Anwendungsinstanz teilen, indem jedem Mandanten ein dedizierter und isolierter Teilbereich einer gemeinsamen Hard- und Softwareinfrastruktur zur Verfügung gestellt wird [55, S. 551] [105, S. 427]. Der Begriff *Mandant* (engl. *tenant*) steht dabei für die einzelnen Dienstnehmer der Software. Im Fall von RTaaS handelt es sich bei den Mandanten also um die jeweiligen Onlineshops oder Bewertungsportale, welche den Dienst in Anspruch nehmen.

¹<http://www.salesforce.com> (besucht am 06.10.2014)

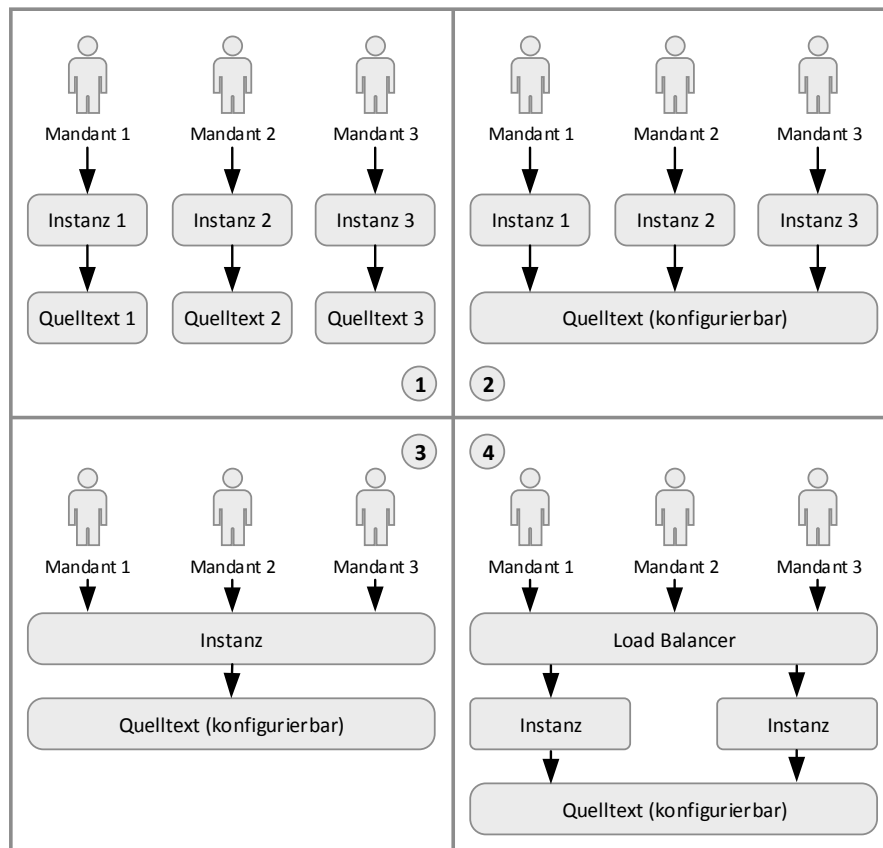


Abbildung 8.2: Vier Ebenen des SaaS-Reifegradmodells (basierend auf [108, S. 180])

Im Zusammenhang mit der Mandantenfähigkeit definieren Chong und Carraro sowie Kwok et al. diverse Anforderungen an eine SaaS [31, 108], welche ebenfalls auf RTaaS übertragbar sind und daher nachfolgend diskutiert werden. In Anlehnung an Chong und Carraro muss ein Dienst im Sinne von SaaS die Anforderungen Skalierbarkeit, Mandanteneffizienz und Konfigurierbarkeit erfüllen [31]. Der Grad der Erfüllung dieser drei Anforderungen bestimmt den *Reifegrad* einer SaaS. Ein *Reifegradmodell* (engl. *maturity model*) besteht aus verschiedenen Stufen hinsichtlich der Reife einer Klasse von Objekten, wobei die unterste Stufe einen initialen Zustand aufweist, während die oberste Stufe für eine vollständige Ausschöpfung des Leistungsvermögens steht [12, S. 213]. Im Hinblick auf SaaS identifizieren Chong und Carraro sowie Kwok et al. vier Reifegrade [31, 108], welche in Abbildung 8.2 aufsteigend nach Reifegrad abgebildet sind.

Auf unterster Ebene des SaaS-Reifegradmodells wird jedem Mandanten eine eigene Anwendungsinstanz bereitgestellt, wobei jede Instanz eine eigene Quelltextbasis besitzt. Bei der Erweiterung dieses Modells (Reifegrad 2) werden jedem Mandanten weiterhin eigene Anwendungsinstanzen zur Verfügung gestellt, welche allerdings jetzt von einer gemeinsamen Quelltextbasis stammen und mandantenspezifisch konfigurierbar sind. In der nächsten Ebene des SaaS-Reifegradmodells werden die mandantenspezifischen Anwen-

dungsinstanzen konsolidiert, d.h., es existiert nur noch eine gemeinsame Instanz, welche von allen Mandanten genutzt wird. Der jeweilige Zugriff wird durch mandantenspezifische Metadaten geregelt. In der höchsten Ebene des SaaS-Reifegradmodells wird zusätzlich auf den Aspekt der Skalierbarkeit fokussiert. Hierzu werden mehrere Instanzen der Anwendung bereitgestellt, welche alle aus einer gemeinsamen Quelltextbasis stammen. Die Zugriffe der Mandanten finden dabei über einen zentralen Load Balancer statt, welcher anhand der aktuellen Auslastung oder anderer Kriterien die Anfragen an eine passende Instanz weiterleitet. Da der RTaaS-Dienst für eine Vielzahl von Dienstnehmer konzipiert ist, welche wiederum eine Vielzahl an Endanwendern besitzen können, wird der in Abbildung 8.2 dargestellte Reifegrad 4 angestrebt, um eine möglichst hohe Skalierbarkeit des Dienstes zu erreichen.

8.3.2 Datenhaltung

Wie aus Abbildung 8.2 hervorgeht, ist eine Mandanteneffizienz im Rahmen des SaaS-Reifegradmodells dadurch gekennzeichnet, dass sich alle Mandanten eine gemeinsame Anwendungsinstanz teilen. In Hinsicht auf die Datenhaltung lässt sich ebenfalls eine Ressourcenteilung vornehmen, um die Mandanteneffizienz zu steigern. Diesbezüglich haben Wang et al. die folgenden Entwurfsmuster zur Datenisolierung innerhalb von Mandantensystemen identifiziert [209, S. 95]:

- **Vollständige Isolation:**
Jeder Mandant erhält eine eigene Datenbank.
- **Teilweise Aufteilung:**
Die Mandanten teilen sich eine gemeinsame Datenbank, allerdings erhält jeder Mandant eigene Tabellen bzw. Schemas.
- **Vollständige Aufteilung:**
Die Mandanten teilen sich sowohl eine gemeinsame Datenbank als auch die darin befindlichen Tabellen bzw. Schemas.

Der Ansatz der vollständigen Isolation weist die geringste Effizienz auf und verursacht im Vergleich zu den anderen oben genannten Ansätzen meist die höchsten Infrastrukturkosten [31]. Es gibt jedoch Anwendungsfälle, z.B. im Bankensektor oder im medizinischen Bereich, welche hohe Anforderungen hinsichtlich der Isolation von Datenbeständen aufweisen und welche daher diesen Ansatz verfolgen müssen [31]. Da im Rahmen von RTaaS in dieser Hinsicht keine speziellen Anforderungen an eine strenge Isolation gegeben sind, ist eine teilweise oder vollständige Aufteilung der Tabellen möglich.

Die dienstorientierte Bereitstellung von Rated Tags für mehrere Mandanten erfordert eine Anpassung und Erweiterung des bestehenden ER-Modells aus Abbildung 4.7 (siehe Seite 64). Die Entitäten des Dienstnehmers umfassen unter anderem die Ressourcen (Produkt oder Dienstleistung), die Endanwender sowie die Kategorisierung der Ressourcen (Produktkategorien). Auf Seite des Dienstgebers befinden sich Basisentitäten für die Verwaltung von Rezensionen und Tags. Neu hinzugekommen sind die Entitäten „Tenant“ und „TagClass“, welche nachfolgend näher erläutert werden.

Die Entität „Tenant“ wird für die Einteilung der Dienstnehmer (Mandanten) benötigt, um die Mandanten-spezifischen Daten voneinander abzugrenzen. Wie später in Abschnitt 8.4 diskutiert wird, soll eine Dienstnehmer-übergreifende Nutzung von Taginformationen stattfinden, um potenzielle Synergieeffekte auszunutzen. Eine solche gemeinsame Nutzung von Tags ist allerdings nur zwischen Dienstnehmern sinnvoll, welche die gleiche Sprache nutzen. So wäre die Einbindung von Tags in deutscher Sprache innerhalb eines englischsprachigen Onlineshops nicht zielführend. Daher enthält die Entität „Tenant“ eine Eigenschaft „LanguageCode“, welche die jeweilige Sprache des Dienstnehmers erfasst. Als standardisierten Ansatz empfiehlt sich hier die Nutzung von Sprachkürzeln gemäß ISO 639¹, z.B. „de“ für Deutsch oder „en“ für Englisch. Da die „Tenant“-Entität für die Mandanten-spezifische Verwaltung eine zentrale Rolle einnimmt, besitzt sie Verbindungen zu jeder anderen Entität innerhalb des RTaaS-Systems (vergleiche hierzu Abbildung 8.3).

Ferner wurde das ER-Modell um die Entität „TagClass“ erweitert. Die Entität beschreibt die Klasse eines Tags und bildet damit die Grundlage, um die im vorigen Kapitel genannte Vereinheitlichung von semantisch ähnlichen Tags umzusetzen. Dienstnehmer haben so die Möglichkeit bereits erstellte oder vordefinierte Tags manuell einer Tag-Klasse zuzuweisen. Ein Beispiel wäre die Zuweisung eines Tags „guter Klang“ zu einer Tag-Klasse „Klangqualität“. Der in Kapitel 7 vorgestellte Klassifikator nutzt diese Trainingsdaten dann für die Klassifikation von neu hinzukommenden Tags. Einer Benutzer-generierten Tag-Klasse können mehrere Tags angehören, wobei ein Tag maximal zu einer Tag-Klasse gehören kann. Das angepasste und erweiterte ER-Diagramm für den RTaaS-Dienst ist in Abbildung 8.3 dargestellt.

Aufgrund des in Abschnitt 8.3.1 vorgeschlagenen Bereitstellungsmodells ergeben sich für den Dienst zusätzliche Anforderungen hinsichtlich der Datenhaltung. Von besonderer Bedeutung ist innerhalb von dienstorientierten Systemen die *Skalierbarkeit*, um eine steigende Anzahl an Dienstnehmer angemessen zu unterstützen. Bondi beschreibt Skalierbarkeit als eine erstrebenswerte Eigenschaft eines Prozesses, Systems oder Netzwerks [19, S. 195]. Gemäß Bondi kann Skalierbarkeit wie folgt definiert werden:

„The concept connotes the ability of a system to accommodate an increasing number of elements or objects, to process growing volumes of work gracefully, and/or to be susceptible to enlargement.“ [19, S. 195]

Da der konzipierte RTaaS-Dienst eine Vielzahl an Dienstnehmern unterstützen soll, ist eine nahtlose Anpassung der Leistung bei steigender Belastung notwendig. Grundsätzlich lassen sich zwei Arten der Skalierung unterscheiden, nämlich die *vertikale Skalierung* (*scale up*) und die *horizontale Skalierung* (*scale out*) [133, S. 1]. Vertikal bedeutet in diesem Zusammenhang, dass die Leistung einer Ressource (z.B. eines Datenbankservers) erhöht wird. Dies könnte beispielsweise durch die Zunahme von Speicherkapazitäten oder schnelleren Prozessoren geschehen. Diese Art der Skalierung ist beispielsweise anwendbar, wenn die dritte Ebene des SaaS-Reifegradmodells (siehe Abbildung 8.2)

¹http://www.iso.org/iso/home/standards/language_codes.htm (besucht am 30.05.2015)

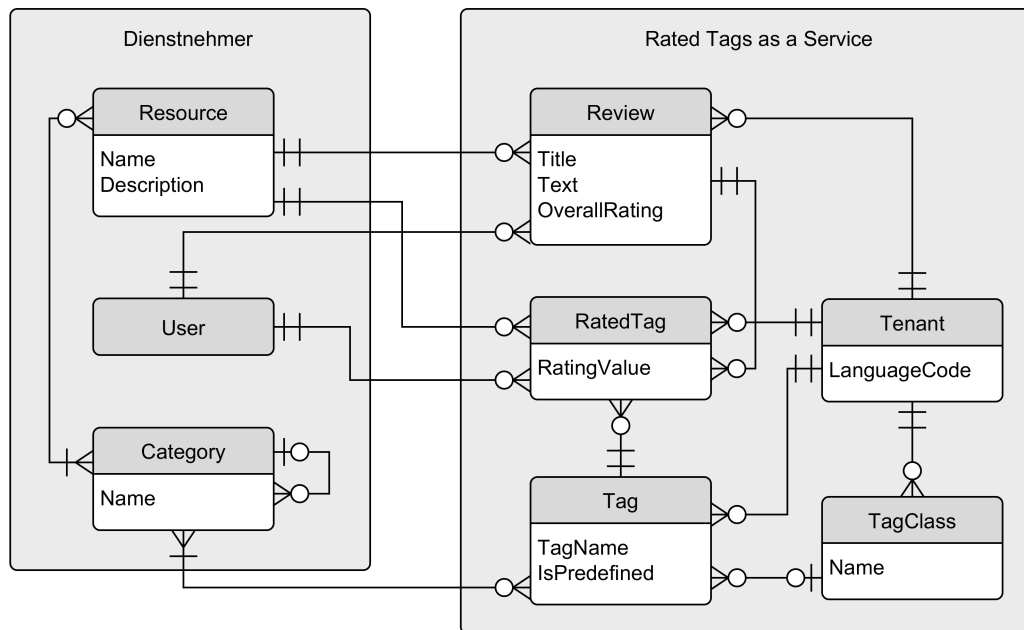


Abbildung 8.3: Angepasstes und erweitertes ER-Diagramm aus Abbildung 4.7 zur Abgrenzung der Entitäten im Rahmen des RTaaS-Dienstes

verwendet wird. Da nur eine Anwendungsinstanz alle Mandanten bedient, wird diese Instanz bei einer stetigen Zunahme an Mandanten immer mehr belastet. Um diese Belastung zu kompensieren, besteht bei diesem Ansatz nur die Möglichkeit die Ressourcen der Anwendungsinstanz (CPU, Speicherplatz etc.) zu erhöhen.

Im Gegensatz zur vertikalen Skalierung steht die horizontale Skalierung für die Verteilung eines Systems auf mehrere miteinander verbundene Systeme, um die eingehende Last zu verteilen. Diese Art der Skalierung ist Reifegrad 4 des SaaS-Reifegradmodells (siehe Abbildung 8.2) zu entnehmen. Im Rahmen der Datenhaltung bedeutet horizontale Skalierung ebenfalls, dass Daten auf verschiedene Systeme repliziert werden müssen. Dies ist nicht nur im Hinblick auf die Skalierung, sondern auch in Hinsicht auf eine höhere Verfügbarkeit der Daten sinnvoll, denn durch die Replikation von Daten wird ein *Single Point of Failure* vermieden. Allerdings muss die Architektur des RTaaS-Dienstes diese Art der Skalierung unterstützen, denn durch die Replikation der Daten auf mehrere Systeme ergeben sich neue Herausforderungen. Diese betreffen vor allem die Aspekte Konsistenz und Verfügbarkeit, welche nachfolgend näher erläutert werden.

Das Ziel einer hohen Verfügbarkeit wird unter anderem durch die Replikation der Daten verfolgt. Dadurch soll erreicht werden, dass der temporäre Ausfall einer Datenbasis die Verfügbarkeit des Systems nicht beeinträchtigt, indem die Daten durch ein anderes Replikat bereitgestellt werden. Dies setzt voraus, dass eine Synchronisation zwischen den Replikaten stattfindet, d.h., Datenänderungen müssen an alle Replikate propagiert werden. Die Art und Weise wie bzw. wann eine Synchronisation durchgeführt wird hängt vom jeweiligen Konsistenzmodell ab. Bei strengen Konsistenzmodellen werden entweder

alle oder keine Replikate verändert, d.h., es wird ein transaktionaler Ansatz verfolgt. Für eine durchgehende Konsistenz der Daten wird hier eine Reduzierung der Verfügbarkeit in Kauf genommen, da die Transaktionsdauer vor allem bei zunehmender Anzahl an Replikaten stark ansteigt und dadurch andere Operationen blockieren kann. Die letztendliche Konsistenz, als Beispiel eines schwachen Konsistenzmodells, garantiert hingegen nur, dass die Daten der Replikate irgendwann (letztendlich) konsistent werden. Es besteht daher die Möglichkeit, dass bei diesem Konsistenzmodell vorübergehend inkonsistente Daten zurückgeliefert werden. Im Gegensatz zu einem strengen Konsistenzmodell ist hier die Verfügbarkeit höher, da Änderungsoperationen verzögert stattfinden und andere Operationen nicht auf die Beendigung einer Transaktion warten müssen.

Die im vorigen Abschnitt aufgezeigte Problematik wird unter anderem durch das *CAP-Theorem* beschrieben. Laut Brewer können in einem verteilten System nur zwei der nachfolgenden Eigenschaften garantiert werden [23]:

- Konsistenz (engl. *consistency*)
- Verfügbarkeit (engl. *availability*)
- Partitionstoleranz (engl. *partition tolerance*)

Neben der Konsistenz und Verfügbarkeit spielt hier auch die Partitionstoleranz, d.h., die Funktionsfähigkeit eines Systems bei voneinander getrennten (partitionierten) Teilsystemen, eine Rolle. Gemäß Vogels muss diese Partitionstoleranz in stark verteilten Systemen, und damit auch im Rahmen von RTaaS, als gegeben angesehen werden [202], wodurch gemäß dem CAP-Theorem nur eine Abwägung zwischen Konsistenz und Verfügbarkeit infrage kommt.

Im Rahmen des RTaaS-Dienstes muss zunächst die Frage geklärt werden, ob vorübergehend inkonsistente Daten in Kauf genommen werden können oder ob ein transaktionaler Ansatz erforderlich ist. Es ist davon auszugehen, dass die Zahl der Leseoperationen die Zahl der Schreiboperationen übersteigt. Da jeder Rezensent nur seine eigene Rezension und die damit verknüpften Rated Tags bearbeitet, ist ein transaktionaler Ansatz hier nicht zwingend erforderlich. Auch die Konsequenzen von temporär inkonsistenten Daten sind nicht als kritisch anzusehen. Aus diesen Gründen kann für den RTaaS-Dienst auf ein strenges Konsistenzmodell verzichtet werden. Für eine Verbesserung der Skalierbarkeit und Verfügbarkeit wird für den RTaaS-Dienst daher ein Modell mit letztendlicher Konsistenz vorgeschlagen.

Ein beispielhaftes UML-Sequenzdiagramm zur Verdeutlichung einer letztendlichen Konsistenz innerhalb des RTaaS-Systems ist in Abbildung 8.4 dargestellt. Es zeigt zwei Clients, zwei Replikate sowie die Zeitpunkte z_1 bis z_4 zu denen jeweils verschiedene Operationen (Serviceanfragen) ausgeführt werden. Eine Änderung von Client 1 mittels Schreiboperation $W(x, t)$ wird zum Zeitpunkt z_1 auf Replikat 1 durchgeführt, jedoch noch nicht weiter propagiert. Eine Leseoperation $R(x)$ auf die veränderte Ressource durch Client 2 würde zum Zeitpunkt z_2 für Replikat 2 noch keinen Wert (*NIL*) zurückliefern, wodurch es hier zu einer temporären Inkonsistenz im Lesezugriff kommt. Erst nach der Propagierung in Zeitpunkt z_3 würde die Leseoperation auf Replikat 2 den Wert t zurückliefern.

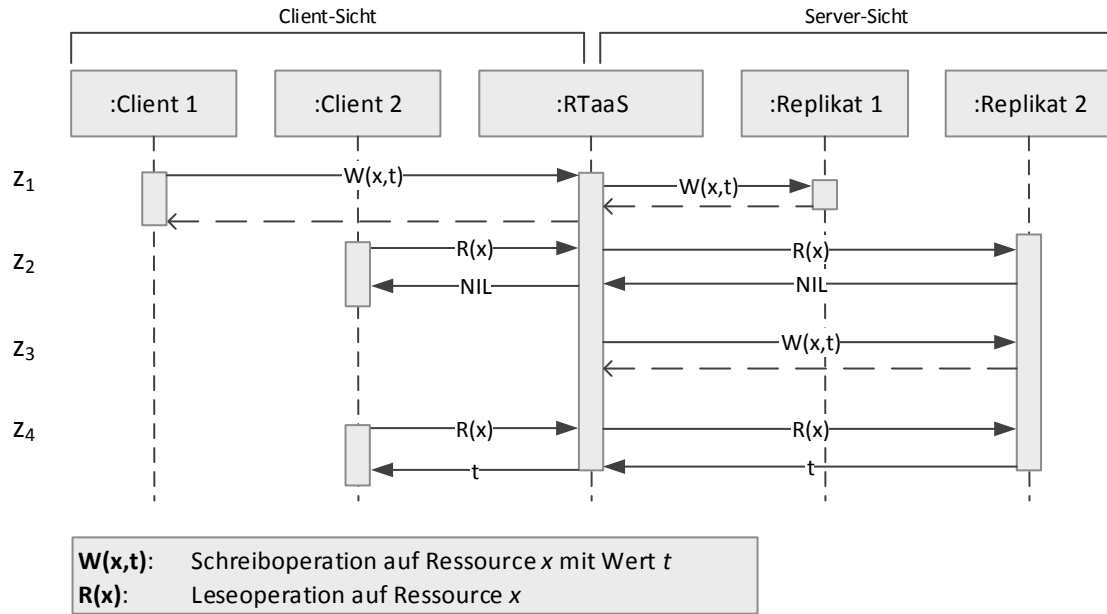


Abbildung 8.4: Beispielhaftes Sequenzdiagramm einer letztendlichen Konsistenz innerhalb von RTaaS (leichte Modifikation von [89, S. 72])

8.3.3 Schnittstellenkommunikation

Für die Nutzung von RTaaS durch Dienstnehmer ist eine Beschreibung der Dienst-schnittstelle erforderlich. Hierzu sind zunächst die für die Kommunikation relevanten Entitäten zu definieren. Diese sind in Quelltext 8.1 dargestellt und entsprechen in leicht abgeänderter Form den Entitäten des Rated Tags-Modells (siehe Kapitel 4). Wie in der *Review*-Entität zu erkennen ist, umfasst diese neben den Rated Tags auch eindeutige Identifikatoren zu der jeweiligen Ressource (*ResourceIdentifier*) und dem Ersteller des Tags (*UserIdentifier*). Dies ist notwendig, da die Ressourcen (d.h. Produkte) und die Endanwender in den Systemen des Dienstnehmers verwaltet werden (vgl. Abbildung 8.3). Um eine Verknüpfung zwischen den Daten beider Systeme herzustellen, sind daher eindeutige Identifikatoren erforderlich. Da der RTaaS-Dienst keinerlei Vorgaben bzgl. des Formats dieser Identifikatoren macht, soll ein Text-basiertes Format verwendet werden. Dabei spielt es keine Rolle, ob Dienstnehmer beispielsweise eine Ganzzahl oder einen *Globally Unique Identifier (GUID)* als Identifikator verwenden, solange eine Konvertierung in Textform möglich ist.

Quelltext 8.1: Entitäten für die Kommunikation mit dem RTaaS-Dienstes in CORBA IDL-Notation

```
module RTaaS.Entities
{
    struct Tag
    {
```

```

    string Identifier;
    string TagName;
    boolean IsPredefined;
    string TagClass;
};

struct TagFilter
{
    string TagIdentifier;
    short MinRating;
    short MaxRating;
};

struct RatedTag
{
    Tag Tag;
    short Rating;
};

struct RatedTagAggregate
{
    Tag Tag;
    float AverageRating;
};

struct Review
{
    string Identifier;
    string Title;
    string Text;
    short OverallRating;
    sequence<RatedTag> RatedTags;
    string ResourceIdentifier;
    string UserIdentifier;
};

struct Category
{
    string Identifier;
    string Name;
    string ParentIdentifier;
};
};

```

Die Schnittstellen zur Nutzung der Kernfunktionalität von RTaaS sind in Quelltext 8.2 aufgeführt. Hierzu gehören unter anderem CRUD-Operationen, d.h., Schnittstellenmethoden zum Erzeugen, Lesen, Aktualisieren und Löschen von Rezensionen und Rated Tags. In Quelltext 8.2 sind dies die Schnittstellenmethoden *createOrUpdateReview*, *deleteReview*, *getReview* und *getReviews*.

Wie an der Schnittstellenmethode *createOrUpdateReview* zu erkennen ist, muss der Dienstnehmer bei dessen Aufruf auch die Kategorien einer Ressource spezifizieren. Dies ist notwendig, damit später anhand einer Kategorie-Kennung alle zugeordneten Ressourcen

ermittelt werden können (siehe Schnittstellenmethode *filterResources* aus Quelltext 8.2). Ohne eine Übergabe dieser Kategorie-Kennungen wäre dies nicht möglich, da die Kategorien in den Systemen des Dienstnehmers verwaltet werden (vgl. Abbildung 8.3) und im RTaaS-System nicht bekannt sind.

Quelltext 8.2: Schnittstellen des RTaaS-Dienstes in CORBA IDL-Notation

```

module RTaaS.Services
{
    import RTaaS.Entities;

    interface RatedTagsService
    {
        boolean createOrUpdateReview(in Review data,
                                   in sequence<string> categoryIds);
        boolean deleteReview(in string resourceId, in string userId);
        Review getReview(in string resourceId, in string userId);
        sequence<Review> getReviews(in string resourceId,
                                  in sequence<TagFilter> tagFilters);
        sequence<RatedTagAggregate> getAggregates(in string resourceId);
        sequence<string> filterResources(in string categoryId,
                                      in sequence<TagFilter> tagFilters);
        sequence<Tag> getSuggestedTags(in string resourceId);
    };
};

```

Für eine aggregierte Darstellung der Rated Tags einer Ressource kann die Schnittstellenmethode *getAggregates* verwendet werden, z.B. um die durchschnittlichen Bewertungen der Rated Tags übersichtlich darzustellen. Zu jedem Tag wird dabei zusätzlich die zugehörige Tag-Klasse (falls vorhanden) übermittelt, wodurch der Dienstnehmer in der Lage ist, Tags an der Oberfläche nach Tag-Klassen zu gruppieren oder diese einzeln darzustellen.

Mithilfe der Schnittstellenmethode *filterResources* sollen Dienstnehmer anhand von Filterbedingungen Ressourcen zu einer bestimmten Kategorie ermitteln können. Dies dient beispielsweise der Reduktion von Produktalternativen durch die Spezifikation von Bewertungsgrenzwerten, z.B. die Ermittlung aller Ressourcen, welche für ein oder mehrere Tags eine definierte durchschnittliche Bewertung übersteigen. Zurückgeliefert werden die Ressourcen-Kennungen, für welche die Bedingungen zutreffen. Der Dienstnehmer ist dann in der Lage die Ressourcen entsprechend zu filtern und anzuzeigen.

Die Schnittstellenmethode *getSuggestedTags* dient der Identifikation von relevanten Tag-Vorschlägen, die einem Anwender bei der Erzeugung von neuen Tags angezeigt werden. Die Methode ermittelt anhand der übergebenen Ressourcen-Kennung zunächst die Kategorien der Ressource. Danach werden alle Tags ermittelt, welche für Ressourcen dieser Kategorien verwendet wurden. Ferner werden die vom Dienstnehmer für die jeweilige Kategorien vordefinierten Tags ausgelesen. Die Vereinigungsmenge dieser beiden Tag-Mengen wird dann an den Dienstnehmer zurückgeliefert.

Zur Verdeutlichung einiger Schnittstellen aus Quelltext 8.2 ist in Abbildung 8.5 ein UML-Sequenzdiagramm mit drei Dienstaufrufen dargestellt. Zunächst wird eine Ressource (Produktseite) aufgerufen, zu der alle zugehörigen Rezensionen samt Rated Tags

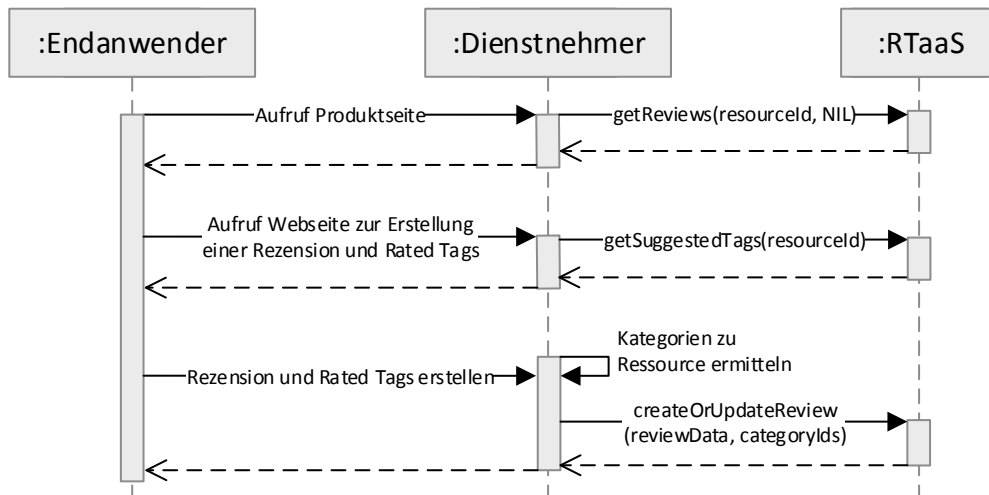


Abbildung 8.5: Beispielhaftes UML-Sequenzdiagramm zur Verdeutlichung einiger Dienstschnittstellen

ausgelesen werden (*getReviews*). Die zweite Aktion des Endanwenders ist der Aufruf einer Webseite zur Erstellung einer neuen Rezension bzw. Rated Tags. Hierzu wird die Schnittstellenmethode *getSuggestedTags* genutzt, um dem Anwender bereits bestehende Tags vorzuschlagen. Die letzte Aktion besteht darin die erstellte Rezension samt Rated Tags zu speichern. Dabei werden beim Dienstnehmer zunächst die Kategorien der jeweiligen Ressource ermittelt, welche später zusammen mit den Daten des Endanwenders an die Schnittstellenmethode *createOrUpdateReview* weitergeleitet werden.

Damit die Dienstnehmer in der Lage sind Tags für bestimmte Kategorien vorzudefinieren sowie Tag-Klassen zu erzeugen und Tags zuzuweisen, wird ein Web Interface innerhalb des RTaaS-Systems vorgeschlagen. Für die Vordefinition und Klassifikation von Tags werden allerdings externe Daten zu den Kategorien benötigt. Diese sind in den Systemen der Dienstnehmer hinterlegt und müssen über eine Dienstschnittstelle dem RTaaS-Dienst zur Verfügung gestellt werden. Die Schnittstelle eines solchen Dienstes ist in Quelltext 8.3 dargestellt. Das RTaaS-System ist dadurch in der Lage die Kategorien der jeweiligen Dienstnehmer zu importieren und innerhalb des Web Interfaces zu verwenden.

Quelltext 8.3: Schnittstelle des Dienstnehmers zur Bereitstellung von Kategorien in CORBA IDL-Notation

```

module Client.Services
{
    import RTaaS.Entities;

    interface CategoryService
    {
        sequence<Category> getCategories();
    };
};
    
```

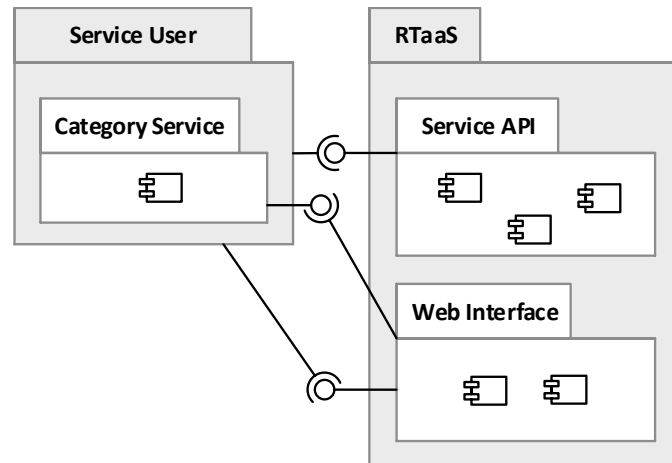


Abbildung 8.6: Abstrakte Darstellung der Schnittstellenkommunikation in Form eines UML-Paketdiagramms

Der Aufbau der oben genannten Kategorieinformationen folgt der in Quelltext 8.1 dargestellten Struktur *Category*. Durch die Möglichkeit zur Angabe von Kategoriehierarchien (*ParentIdentifier*) verringert sich die Redundanz im Rahmen der Vordefinition von Tags. Dies soll am Beispiel einer Kategorie „Kopfhörer“ mit den Unterkategorien „Schnurlose Kopfhörer“ und „Kabelgebundene Kopfhörer“ verdeutlicht werden. Allgemeine Tags, z.B. „Klangqualität“, lassen sich der Oberkategorie zuordnen und sind dann auch automatisch für Unterkategorien nutzbar. Eine erneute Vordefinition dieser Tags ist in den Unterkategorien nicht mehr erforderlich.

Die Vordefinition und Klassifikation von Tags kann die Qualität des Systems steigern, jedoch ist deren Nutzung optional. Werden diese Funktionalitäten nicht benötigt, so kann auf die in Quelltext 8.3 dargestellte Schnittstelle verzichtet werden. Jedoch ist dann eine teilautomatisierte Vereinheitlichung der Tags (wie in Kapitel 7 beschrieben) nicht mehr möglich.

Es ist zu beachten, dass die oben genannten Ausführungen die Kernfunktionalitäten von RTaaS beschreiben und keine allumfassende Implementierungsspezifikation darstellen. Zusätzliche Funktionalitäten der RTaaS-Schnittstellen sind für zukünftige Implementierungen denkbar, z.B. eine Volltextsuche für Rezensionen. Ferner ist im Rahmen der Schnittstellenkommunikation die jeweilige Mandantenkennung zu übermitteln, um die Transaktionen den jeweiligen Dienstnehmern zuzuordnen. Diese Mandantenkennung wurde aus Gründen der Übersichtlichkeit nicht in den genannten Schnittstellenmethoden aus Quelltext 8.2 aufgeführt. Abschließend soll Abbildung 8.6 einen abstrakten Überblick zu den Komponenten der Schnittstellenkommunikation liefern. Die Abbildung zeigt die Abhängigkeiten zwischen den Kommunikationsteilnehmern. Der RTaaS-Dienst stellt über die Service API und das Web Interface Funktionalitäten bereit, während das Web Interface eine Abhängigkeit zum *Category Service* des Dienstnehmers besitzt, um die dort hinterlegten Kategorien zu nutzen.

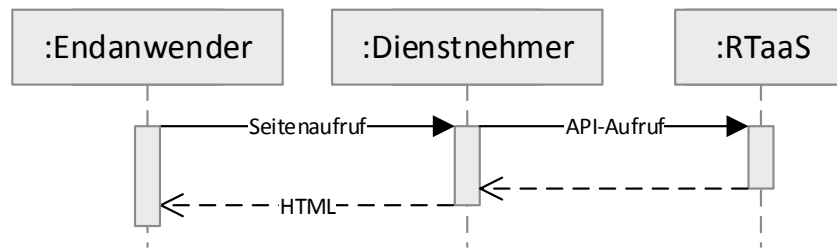


Abbildung 8.7: UML-Sequenzdiagramm einer synchronen Integration des RTaaS-Dienstes

8.3.4 Dienststeinbindung

Basierend auf der in vorigen Abschnitt diskutierten Schnittstellenkommunikation, soll nun die Einbindung des Dienstes aus Sicht des Dienstnehmers betrachtet werden. Wie in Abbildung 8.1 bereits veranschaulicht wurde, findet eine direkte Kommunikation zwischen Dienstnehmer und RTaaS-Dienst statt. Der Endanwender des Dienstes nimmt von der Existenz des RTaaS-Dienstes keine Kenntnis. Dies unterscheidet sich beispielsweise vom Kommentar-Service Disqus¹, welcher von einigen Webseiten zur Einbindung von Kommentarfunktionalität genutzt wird. Dort benötigt jeder Endanwender ein eigenes Konto bei dem Dienstgeber Disqus.

Für die Einbindung des RTaaS-Dienstes durch Dienstnehmer ergeben sich generell zwei Alternativen. Zum einen ist ein synchroner Aufruf des Dienstes möglich, bei der die vollständige Auslieferung bzw. Darstellung einer Webseite so lange blockiert wird, bis die Antwort des Dienstaufwurfes eingetroffen ist. Ein solcher Ansatz ist in Abbildung 8.7 dargestellt. Ein Endanwender fordert eine Webseite an, welche wiederum Informationen des RTaaS-Dienstes benötigt.

Die zweite Alternative zur Dienststeinbindung verwendet einen asynchronen Ansatz. Hier werden alle benötigten Seiteninhalte ohne vorige Kommunikation mit dem RTaaS-Dienst zurückgeliefert. Innerhalb des zurückgelieferten HTML-Quelltextes wird JavaScript-Quelltext hinterlegt, welcher beim Laden der Webseite das asynchrone Nachladen von Inhalten vornimmt. Mit Hilfe von AJAX wird ein Endpunkt des Dienstnehmers angesprochen, welcher sich wiederum Inhalte über die API des RTaaS-Dienstes beschafft. Dieses Vorgehen wird in Abbildung 8.8 veranschaulicht.

Die Wahl einer der oben genannten Alternativen ist letztlich von den Anforderungen des Dienstnehmers abhängig. Muss beispielsweise auf den Einsatz von JavaScript verzichtet werden, so bleibt nur eine synchrone Integration des RTaaS-Dienstes. Dies wiederum kann sich negativ auf die Reaktionsfähigkeit der Anwendung auswirken, da die Generierung und Auslieferung der HTML-Seite verzögert wird. Dieser Nachteil wird im asynchronen Ansatz umgangen, indem die Inhalte nachgeladen werden und somit die Reaktionsfähigkeit der Anwendung nur geringfügig beeinträchtigt wird.

¹<http://disqus.com> (besucht am 17.06.2015)

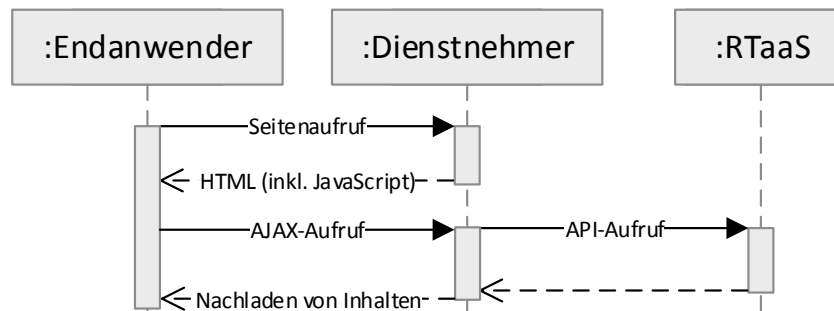


Abbildung 8.8: UML-Sequenzdiagramm einer asynchronen Integration des RTaaS-Dienstes durch das Nachladen von Inhalten mittels JavaScript und AJAX

8.4 Dienstnehmer-übergreifende Nutzung aller Tags

Wie bereits in Abschnitt 8.1 erwähnt wurde, soll an dieser Stelle ein Modell vorgestellt werden, um die Benutzer-generierten Tags aller Dienstnehmer gemeinsam nutzbar zu machen. Das Ziel dieses Ansatzes ist, dass für jeden Dienstnehmer nicht nur die Benutzer-generierten Tags der eigenen Anwender, sondern auch die Tags der Anwender anderen Dienstnehmer sichtbar sind und so ein besserer Überblick über die relevanten Produktaspekte einer Produktgruppe entsteht. Ferner soll durch die Kombination der Tags die Kaltstartproblematik des einzelnen Dienstnehmers reduziert werden. Die Nutzung von Tags anderer Dienstnehmer ist allerdings nur dann sinnvoll, wenn diese ebenfalls in der gleichen Sprache definiert wurden. Dies kann über das Sprachkürzel des jeweiligen Mandanten sichergestellt werden (vgl. „LanguageCode“ aus Abbildung 8.3).

Ein zentrales Problem der Dienstnehmer-übergreifende Nutzung aller Tags ist, dass eine Verbindung zwischen den Ressourcen der einzelnen Dienstnehmern hergestellt werden muss, d.h., es muss eine einheitliche Kategorisierung der Ressourcen möglich sein. Zunächst stellt sich die Frage wie identische Produkte identifiziert werden können, welche von unterschiedlichen Dienstnehmern verwaltet werden. Hierzu bietet sich die Nutzung der Global Trade Item Number (GTIN)¹ an, welche eine Nummer zur eindeutigen Identifikation von Produkten darstellt. Produkte im E-Commerce besitzen häufig eine GTIN bzw. ein dazu kompatibles Format, z.B. eine ISBN. Eine GTIN ist allerdings nur für Produkte und nicht für Dienstleistungen anwendbar. Jeder Dienstnehmer muss also zu einem Produkt die jeweilige GTIN bereitstellen. Dadurch können identische Produkte und somit auch deren erstellte Tags, Dienstnehmer-übergreifend ermittelt werden. Die Menge der relevanten Tags ergibt sich dann aus der Vereinigung der einzelnen Tag-Mengen für eine GTIN.

Der oben genannte Ansatz bezieht sich auf die Produktebene. Ein größerer Mehrwert hinsichtlich der Dienstnehmer-übergreifenden Nutzung aller Tags wäre jedoch

¹<http://www.gs1.org/gtin> (besucht am 22.06.2015)

die Adressierung der Produktklassenebene. Dies würde bedeuten, dass auch Tags aus gleichen Produktklassen übergreifend genutzt werden können und nicht nur Tags aus Produkten mit gleicher GTIN. Die Ermittlung der jeweiligen Klasse eines Produkts gestaltet sich allerdings problematisch, da kein einheitliches System zur Produktklassifikation vorhanden ist. Einige standardisierte Ansätze umfassen eClass¹, UNSPSC² oder RosettaNet³, wobei jedoch auch häufig unternehmenseigene Klassifikationssysteme zum Einsatz kommen. Es kann also nicht davon ausgegangen werden, dass Dienstnehmer eine einheitliche Methodik zur Produktklassifikation verwenden. Als möglicher Ansatz wird daher die Nutzung der ProductOntology⁴ von Martin Hepp vorgeschlagen. Es handelt sich dabei um eine leichtgewichtige Ontologie, welche speziell für die Produkt- oder Dienstklassifikation im E-Commerce entwickelt wurde. Sie basiert auf der Extraktion von Informationen aus Wikipedia⁵ zur Abbildung der Ontologie [67]. Jedes Element der Ontologie ist über einen Uniform Resource Identifier (URI) eindeutig definiert. Produkte einer Klasse „Kopfhörer“ wären beispielsweise über den URI <http://www.productontology.org/id/Headphones> identifizierbar.

Die Nutzung von ProductOntology hat gegenüber den oben genannten Produktklassifikationssystemen den Vorteil, dass die Klassifikation nicht nur für Produkte, sondern auch für sonstige Ressourcen (z.B. Dienstleistungen) ermöglicht wird. Die geringe Komplexität der leichtgewichtigen Ontologie vereinfacht die Anwendung, wobei jedoch im Vergleich zu Produktklassifikationssystemen häufig keine Produktklassenhierarchien abbildbar sind. Für das oben genannte Element „Kopfhörer“ sind beispielsweise keine Unterkategorien, wie z.B. „Funk-Kopfhörer“ abbildbar. Dies liegt daran, dass für Funk-Kopfhörer kein eigener URI existiert.

Eine beispielhafte Zuweisung eines Elements der ProductOntology zu einer Produktklasse eines Dienstnehmers ist in Abbildung 8.9 dargestellt. Die Abbildung zeigt, dass seitens des Dienstnehmers eine Zuordnung definiert werden muss, welche einer Produktkategorie ein Element der ProductOntology zuweist. Dies könnte beispielsweise über das in Abschnitt 8.3.3 beschriebene Web Interface vorgenommen werden. Entsprechend muss dem bestehenden Datenmodell eine neue Entität hinzugefügt werden, welche die Elemente der ProductOntology verwaltet (siehe Entität „ProductOntology“ aus Abbildung 8.10). Zwischen den Entitäten „Category“ und „ProductOntology“ besteht eine 1:1-Beziehung, die der Zuweisung aus Abbildung 8.9 entspricht. Ferner verdeutlicht die Beziehung zwischen den Entitäten „Tenant“ und „ProductOntology“, dass eine konkrete Ausprägung der Entität „ProductOntology“ von einem oder mehreren Mandanten genutzt werden kann.

Die beiden Möglichkeiten zur Dienstnehmer-übergreifenden Nutzung von Tags sind in Abbildung 8.11 dargestellt. Die übergreifenden Tags lassen sich von den Dienstnehmern

¹<http://www.eclass.de> (besucht am 22.06.2015)

²<http://www.unspsc.org> (besucht am 22.06.2015)

³<http://www.rosettanet.org> (besucht am 22.06.2015)

⁴<http://www.productontology.org> (besucht am 22.06.2015)

⁵<http://www.wikipedia.org> (besucht am 22.06.2015)

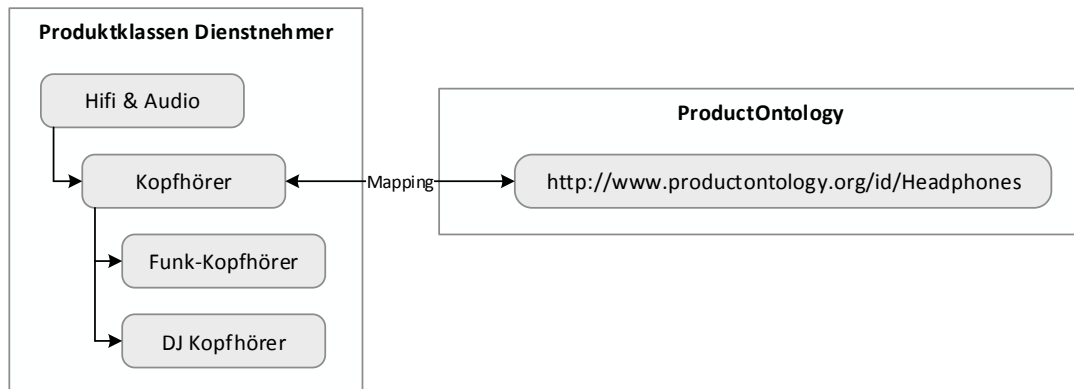


Abbildung 8.9: Beispielhafte Zuweisung eines Elements der ProductOntology an eine Produktklasse des Dienstnehmers

integrieren und liefern so einen erweiterten Überblick zu den diskutierten Produktaspekten. Zudem erhöht sich dadurch auch die Anzahl der potenziellen Trainingsdaten, welche für eine teilautomatisierte Vereinheitlichung der Tags verwendet werden können (vgl. Abschnitt 8.4). Es ist abschließend anzumerken, dass eine Wiederverwendung eines Klassifikationsschemas (Tag-Klasse) anderer Dienstnehmer nicht möglich bzw. sinnvoll erscheint. Dies liegt daran, dass je nach Dienstnehmer andere Anforderungen hinsichtlich der Klassifikation vorliegen. Während ein Dienstnehmer einen Tag „Bildschärfe“ beispielsweise als eigenständige Klasse definiert, ordnet ein anderer Dienstnehmer diesen Tag beispielsweise der Klasse „Bildqualität“ zu. Eine Dienstnehmer-übergreifende Nutzung eines Klassifikationsschemas ist daher nicht zielführend.

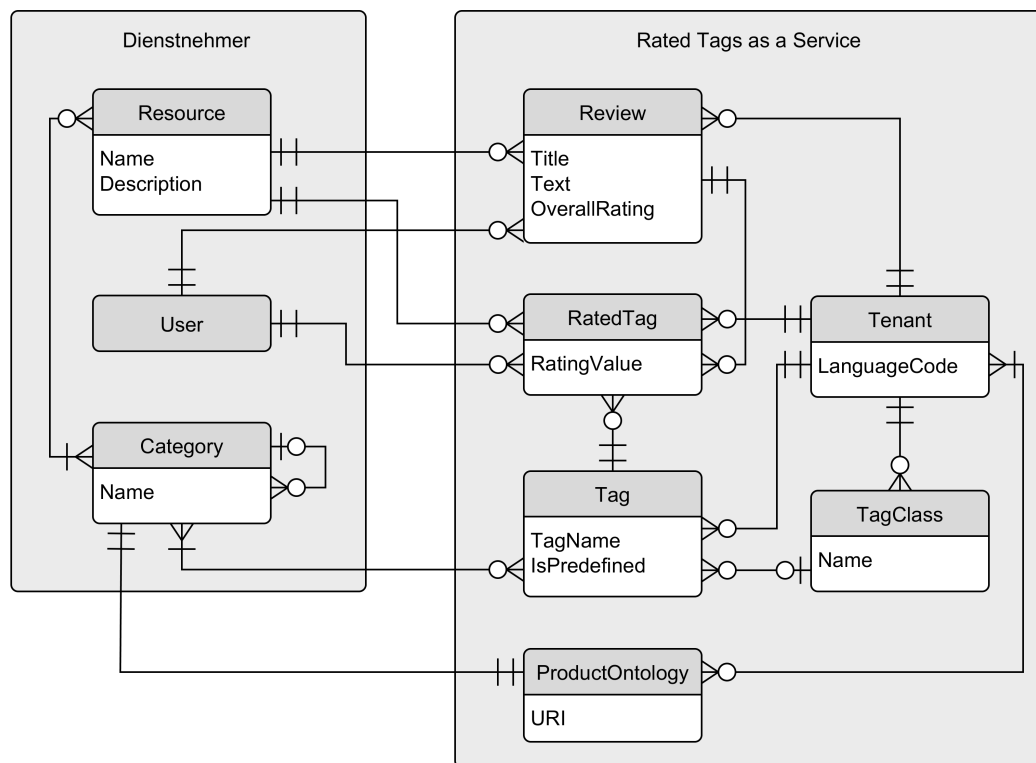


Abbildung 8.10: Erweitertes ER-Diagramm aus Abbildung 8.3 zur Unterstützung einer Dienstnehmer-übergreifenden Nutzung aller Tags

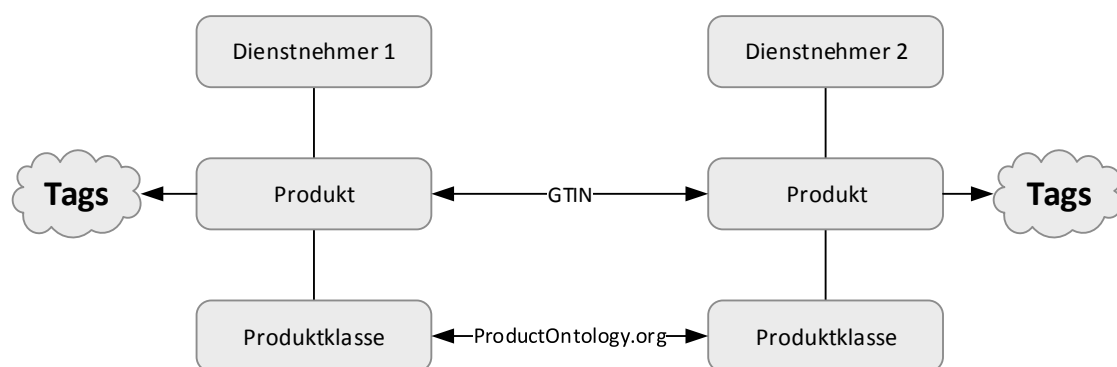


Abbildung 8.11: Teilen von Tag-Informationen einzelner Produkte anhand von GTIN und der Produktklassenontologie ProductOntology.org

Kapitel 9

Zusammenfassung

9.1 Ergebnisse der Arbeit

Die vorliegende Arbeit hat sich mit der Konzeptionierung, Evaluierung und Dienstorientierten Bereitstellung einer interaktiven Entscheidungshilfe für den E-Commerce beschäftigt. Hierzu wurden in den Kapiteln 2 und 3 zunächst die Grundlagen und verwandten Arbeiten aus den Bereichen E-Commerce und Social Media analysiert. In Abschnitt 3.3 wurde die Durchführung einer empirischen Studie dokumentiert, welche den Einsatz von Social Media in den umsatzstärksten B2C-Onlineshops aus Deutschland untersucht hat. Die Studie liefert somit einen Beitrag für ein besseres Verständnis zum Stand der Technik im Forschungsbereich des Social Commerce. In Bezug auf die sieben Bausteine des Social Media wurde im Rahmen der Studie ermittelt, dass vor allem der Austausch von Inhalten sowie der Baustein „Reputation“ in aktuellen Onlineshops große Anwendung finden. Die beiden am häufigsten verwendeten Social Media Features waren Produktbewertungen und Kundenrezensionen. Es hat sich gezeigt, dass Onlineshops diese beiden Funktionalitäten fast immer gemeinsam verwenden.

In Anlehnung an die in der empirischen Studie ermittelten Ergebnisse wurden die Elemente Produktbewertung und Kundenrezension näher betrachtet. Die Forschungsarbeit von Vig et al. hat diesbezüglich gezeigt, dass in bestehenden Bewertungssystemen eine Lücke zwischen diesen beiden Ansätzen existiert [200, S. 323]. Bezug nehmend dazu wurde daher ein Konzept entworfen, welches diese Lücke schließen und somit die Entscheidungsfindung von Konsumenten erleichtern soll. Für das entworfene Konzept mit der Bezeichnung *Rated Tags* wurden Ansätze aus den Bereichen Social Tagging und Bewertungssysteme kombiniert. Der Rated Tags-Ansatz kann als interaktive Internet-basierte Entscheidungshilfe bezeichnet und somit der Kategorie der interaktiven Informationsverwaltungs-Tools (IIMTs) zugeordnet werden. Die Ziele von Rated Tags sind die Erhöhung der Entscheidungsqualität sowie die Reduzierung des Entscheidungsaufwands für Konsumenten.

In Kapitel 4 wurde ferner die konzeptionelle Betrachtung sowie die Abgrenzung des Rated Tags-Ansatzes zu verwandten Arbeiten diskutiert. Als Ergebnis der Abgrenzung lässt sich festhalten, dass bereits verwandte Forschungsarbeiten existieren, welche die Elemente Tagging und Bewertung miteinander kombinieren, allerdings ist die Kontextspezifische Überführung in eine interaktive Entscheidungshilfe ein Alleinstellungsmerkmal des Rated Tags-Ansatzes. Ausgehend von dem fachlichen Konzept wurden in Kapitel 5 dann die Anforderungen und Umsetzungsmöglichkeiten aus technischer Sicht betrachtet. Dabei wurden die Herausforderungen und Anforderungen der Systemarchitektur wie

auch der Benutzeroberfläche näher analysiert. Weiterhin wurde eine prototypische Implementierung des Rated Tags-Ansatzes präsentiert, um die Tauglichkeit des aufgestellten Modells zu erproben. Im Hinblick auf eine spätere Evaluierung wurde für den Prototyp eine Web-basierte Implementierungsvariante gewählt. Unter Anwendung von dynamischen Ansätzen im Frontend-Bereich (wie beispielsweise AJAX), konnte insbesondere der interaktive Charakter der Entscheidungshilfe gut umgesetzt werden, z.B. die dynamische Aktivierung und Deaktivierung eines Filters. Die Anwendung von Entwurfsmustern, wie etwa dem MVC-Muster, hat ferner demonstriert, wie eine modulare Umsetzung des entworfenen Konzepts aussehen könnte.

Basierend auf dem erstellten Prototyp wurde in Kapitel 6 eine Anwenderstudie konzipiert und durchgeführt, um die Praxistauglichkeit und Benutzerakzeptanz des Rated Tags-Ansatzes zu evaluieren. Hierfür wurden als Datenbasis reale Kundenrezensionen des Onlinehändlers Amazon.de verwendet. Das Hauptziel der Evaluierung war es zu untersuchen, ob Konsumenten durch die Verwendung von Rated Tags eine verbesserte Entscheidungsqualität sowie einen reduzierten Entscheidungsaufwand aufweisen. Für diese Überprüfung wurden entsprechende Hypothesen aufgestellt, die anhand eines Kontroll-experiments überprüft wurden. Hierzu wurden insgesamt 34 Studienteilnehmer akquiriert, wovon jeweils die Hälfte der Teilnehmer auf Kontroll- und Experimentalgruppe verteilt wurden. Der Experimentalgruppe (auch Rated Tags-Gruppe bezeichnet) wurde neben den herkömmlichen Rezensionen ebenfalls die Rated Tags-Entscheidungshilfe zur Verfügung gestellt, während die Kontrollgruppe keinerlei Kenntnis über Rated Tags hatte. Die Auswertung der Studienergebnisse hat gezeigt, dass Rated Tags bei den Teilnehmern sowohl die Qualität der Entscheidung gesteigert, als auch den kognitiven Aufwand reduziert haben. Des Weiteren haben die Ergebnisse der Studie gezeigt, dass der Großteil der erstellten Rated Tags eine hohe Qualität aufweist und eine hohe Wiederverwendbarkeit bestehender Tags erkennbar war. Daraus lässt sich schließen, dass Anwender auch ohne eine zentrale Kontrollinstanz einen Konsens in Bezug auf relevante Tags erreichen. Aus den Ergebnissen des Fragebogens konnte zudem ermittelt werden, dass sowohl die Nützlichkeit als auch die Bedienbarkeit des Rated Tags-Konzepts von den Anwendern als gut bis sehr gut eingestuft wurde.

Obwohl im Rahmen der Studie eine hohe Wiederverwendung von Tags erkennbar war, entstanden doch einige Tags mit ähnlicher Semantik, z.B. die Tags „Preis/Leistung“ und „Preis-Leistungsverhältnis“. Die gedankliche Zusammenführung solcher Tags bedeutet für den Konsumenten einen erhöhten kognitiven Aufwand. Um diesen Aufwand zu reduzieren, wurde in Kapitel 7 ein Konzept zur teilautomatisierten Vereinheitlichung von semantisch ähnlichen Tags erarbeitet und evaluiert. Hierfür wurde ein Ensemble-Klassifikator aus dem Bereich des überwachten maschinellen Lernens entworfen, welcher zur Lösung des Klassifikationsproblems unter anderem verschiedene heterogene Basisklassifikatoren verwendet. Die Evaluierung des entworfenen Klassifikators anhand einer großen Menge an realen Datensätzen hat gezeigt, dass die Leistung des Klassifikators den momentanen Stand der Technik hinsichtlich der Klassifikation von kurzen Benutzeraspekten übersteigt. Ferner konnte anhand der analysierten Daten verdeutlicht werden, dass die erstellten und wiederverwendeten Benutzeraspekte Potenzgesetzen folgen, welche typischerweise im Social Tagging anzutreffen sind.

	<u>Produkt #1</u>	<u>Produkt #2</u>	<u>Produkt #3</u>
Bildqualität	3,5 Sterne	3,1 Sterne	4,2 Sterne
Klang	2,3 Sterne	4,4 Sterne	4,0 Sterne
Preis-/Leistung	3,0 Sterne	4,5 Sterne	3,3 Sterne
Design	4,1 Sterne	3,9 Sterne	3,5 Sterne

Abbildung 9.1: Beispielhafte Darstellung einer Vergleichsmatrix unter Verwendung von Rated Tags

Als Abschluss dieser Arbeit wurde in Kapitel 8 ein Konzept vorgestellt, um Rated Tags in Service-orientierter Weise für Bewertungsportale oder Onlineshops zur Verfügung zu stellen. Es wurden die speziellen Anforderungen hinsichtlich der Dienstarchitektur, Datenhaltung sowie Diensteinbindung diskutiert. Hierzu wurde unter anderem das ER-Modell des Rated Tags-Ansatzes erweitert, um die Mandanten-spezifischen Eigenheiten zu berücksichtigen. Ferner wurde ein Modell vorgestellt, um eine Dienstnehmer-übergreifende Nutzung der erstellten Tags zu ermöglichen und potenzielle Synergieeffekte auszunutzen.

9.2 Ausblick

Wie der vorherige Abschnitt verdeutlicht hat, wurden in dieser Arbeit einige vielversprechende Ergebnisse erzielt. Eine weitere Vertiefung des Forschungsthemas sowie eine Anwendung auf verwandte Themenkomplexe wären daher als sinnvoll anzusehen. Aufgrund des interdisziplinären Charakters der Forschungsarbeit kämen für eine zukünftige Forschung einige Anknüpfungspunkte infrage, welche nachfolgend kurz diskutiert werden.

Ein potenzieller Ansatzpunkt für eine zukünftige Forschungsarbeit wäre die Verwendung der Benutzer-generierten Rated Tags innerhalb von bestehenden Entscheidungshilfen. Im Bereich der interaktiven Entscheidungshilfen wäre eine Integration des Rated Tags-Ansatzes innerhalb einer Vergleichstabelle denkbar. Wie bereits in Abschnitt 2.3.3 erläutert wurde, umfasst eine Vergleichstabelle mögliche Produktalternativen und deren Produktmerkmale, wodurch ein direkter Produktvergleich auf Merkmalsebene möglich ist. Die typischerweise objektiven Kriterien einer Vergleichstabelle (z.B. Größe) könnten durch den Einsatz von Rated Tags um subjektive Kriterien erweitert werden, da die erzeugten Tags letztlich auch Produktmerkmale darstellen. Ein früher Prototyp einer solchen Vergleichsmatrix mit integrierten Rated Tags ist in Abbildung 9.1 dargestellt. Die Abbildung zeigt beispielhaft drei Produktalternativen sowie die Benutzer-generierten Rated Tags als Produktmerkmale. Anhand der Hintergrundfarbe einer Zelle könnte beispielsweise das dominierende Produkt hinsichtlich eines Merkmals kenntlich gemacht werden. Eine zukünftige Forschungsarbeit in diesem Bereich müsste die speziellen Herausforderungen des Ansatzes erproben. So ist es beispielsweise denkbar, dass ein Rated

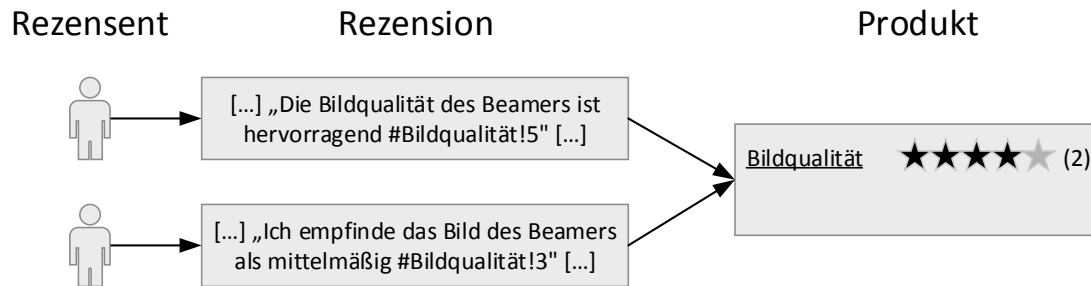


Abbildung 9.2: Beispielhafter Anwendungsfall für den Einsatz von bewertbaren Hash-tags in Rezensionen

Tag nicht für alle dargestellten Produkte vergeben wurde, wodurch ein direkter Vergleich erschwert würde. Eine zusätzliche Erweiterungsmöglichkeit der Vergleichstabelle wäre die Entwicklung einer Filtermöglichkeit, damit den Konsumenten nicht alle, sondern nur die für sie relevanten Merkmale angezeigt werden.

Wie bereits in Abschnitt 6.3.3 erwähnt wurde, ergab sich durch das qualitative Feedback eines Studienteilnehmers eine interessante Erweiterungsidee für den Rated Tags-Ansatz. Die Idee ist, nicht nur Rated Tags zu einer Rezension zu vergeben, sondern auch die entsprechende Textstelle zu kennzeichnen, an der das jeweilige Produktmerkmal diskutiert wird. Dies würde vor allem bei längeren Rezensionen die Suche nach dem diskutierten Aspekt deutlich erleichtern. Basierend auf dieser Problemstellung wurde ein erstes Konzept erarbeitet und prototypisch implementiert sowie ein entsprechender Artikel mit dem Titel „Combining Hashtags and Ratings for Opinion Expression in Customer Reviews“ publiziert [88].

Die Basis für das Konzept stellen Hashtags dar, welche in sozialen Netzwerken wie Twitter vor allem für die Klassifizierung von Gesprächsthemen verwendet werden. Diese normalen Hashtags werden im Rahmen des Konzepts um eine zusätzliche Bewertungskomponente erweitert. Dies bedeutet, dass diese erweiterten Hashtags nicht mehr nur als Inhaltsklassifikator, sondern auch für Bewertungszwecke verwendet werden können. Ein Beispiel für die Anwendung solcher bewertbaren Hashtags ist in Abbildung 9.2 zu finden. Die Abbildung zeigt einen Anwendungsfall, bei dem zwei Rezensenten innerhalb ihrer Rezension diese erweiterte Form der Hashtags (hier „#Bildqualität!5“) verwenden, um ein Produktmerkmal zu bewerten. Wie der Aufbau des obigen Hashtags zeigt, werden das Produktmerkmal („Bildqualität“) und die dazugehörige Bewertung (5) hier beispielhaft durch ein Ausrufezeichen getrennt. Anhand einer Textanalyse der jeweiligen Rezensionen können diese Hashtags samt Bewertung dann später extrahiert und weiterverarbeitet werden. Der Anwendungsfall aus Abbildung 9.2 zeigt beispielsweise die aggregierte Darstellung der beiden Hashtag-Bewertungen eines Produkts am Beispiel des Hashtags „#Bildqualität“.

Im Rahmen einer weiteren Forschungsarbeit müsste das oben dargestellte Konzept so erweitert werden, dass Anwendern eine einfache Wiederverwendung von bereits bestehenden Hashtags ermöglicht wird. Dies ist notwendig um die Qualität zu erhöhen und redundante

Hashtags zu vermeiden. Ferner müsste die Praxistauglichkeit und Benutzerakzeptanz des neuen Konzepts, beispielsweise im Rahmen einer weiteren Anwenderstudie, erprobt werden.

Als abschließender Ausblick auf zukünftige Arbeiten ist die Implementierung und Evaluierung des Service-orientierten Entwurfs von Rated Tags zu nennen. Das in Kapitel 8 vorgestellte Modell kann hierfür als Grundlage verwendet werden. Für die Implementierung des Dienstes ist die Nutzung von Cloud-basierten Plattformen (Platform as a Service) denkbar, welche bereits häufig Automatismen zur Verbesserung der Skalierbarkeit bereitstellen. In einer zukünftigen Evaluierung könnte dann beispielsweise untersucht werden, wie die Einbindung des Dienstes die Performance des Dienstnehmers beeinträchtigt.

Literaturverzeichnis

- [1] Gediminas Adomavicius und Young Ok Kwon. „New Recommendation Techniques for Multicriteria Rating Systems“. In: *IEEE Intelligent Systems* 22(3) (Mai 2007), S. 48–55.
- [2] Charu C. Aggarwal und ChengXiang Zhai. „A Survey of Text Classification Algorithms“. In: *Mining Text Data*. Hrsg. von Charu C. Aggarwal und ChengXiang Zhai. Springer US, 2012, S. 163–222.
- [3] David W. Aha. „Editorial“. In: *Lazy Learning*. Hrsg. von David W. Aha. Norwell, MA, USA: Kluwer Academic Publishers, 1997, S. 7–10.
- [4] James F. Allen und Alan M. Frisch. „What’s in a Semantic Network?“ In: *Proceedings of the 20th Annual Meeting on Association for Computational Linguistics*. Toronto, Ontario, Kanada: Association for Computational Linguistics, 1982, S. 19–27.
- [5] Lloyd Allison und Trevor I. Dix. „A bit-string longest-common-subsequence algorithm“. In: *Information Processing Letters* 23(5) (1986), S. 305–310.
- [6] Morgan Ames und Mor Naaman. „Why We Tag: Motivations for Annotation in Mobile and Online Media“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’07. San Jose, Kalifornien, USA: ACM, 2007, S. 971–980.
- [7] Gad Ariav und Michael J. Ginzberg. „DSS Design: A Systemic View of Decision Support“. In: *Communications of the ACM* 28(10) (Okt. 1985), S. 1045–1052.
- [8] Dan Ariely. „Controlling the information flow: Effects on consumers’ decision making and preferences“. In: *Journal of Consumer Research* 27(2) (2000), S. 233–248.
- [9] Michael Bächle. „Ökonomische Perspektiven des Web 2.0 – Open Innovation, Social Commerce und Enterprise 2.0“. In: *Wirtschaftsinformatik* 50(2) (Juni 2008), S. 129–132.
- [10] Helmut Balzert. *Lehrbuch der Softwaretechnik: Entwurf, Implementierung, Installation und Betrieb*. 3. Auflage. Lehrbücher der Informatik. Spektrum Verlag, 2011.
- [11] Astrid Beck. „Web 2.0: Konzepte, Technologie, Anwendungen“. In: *HMD – Praxis der Wirtschaftsinformatik* 44(255) (Juni 2007), S. 5–16.
- [12] Jörg Becker, Ralf Knackstedt und Jens Pöppelbuß. „Developing Maturity Models for IT Management – A Procedure Model and its Application“. In: *Business & Information Systems Engineering* 1(3) (2009), S. 213–222.
- [13] Izak Benbasat und Barrie R. Nault. „An evaluation of empirical research in managerial support systems“. In: *Decision Support Systems* 6(3) (1990), S. 203–226.

- [14] James R. Bettman, Eric J. Johnson und John W. Payne. „A componential analysis of cognitive effort in choice“. In: *Organizational Behavior and Human Decision Processes* 45(1) (1990), S. 111–139.
- [15] Phil Bianco, Rick Kotermanski und Paulo Merson. *Evaluating a Service-Oriented Architecture*. Techn. Ber. CMU/SEI-2007-TR-015. Software Engineering Institute, Sep. 2007.
- [16] Haji Binali, Vidyasagar Potdar und Chen Wu. „A state of the art opinion mining and its application domains“. In: *Proceedings of the IEEE International Conference on Industrial Technology*. Gippsland, Victoria, Australien, Feb. 2009, S. 1–6.
- [17] BITKOM. *Online-Bewertungen liegen im Trend*. Juni 2013. URL: http://www.bitkom.org/de/presse/78284_76564.aspx (besucht am 05.04.2014).
- [18] BITKOM. *Online-Shopping in Deutschland besonders beliebt*. Okt. 2013. URL: http://www.bitkom.org/de/presse/78284_77580.aspx (besucht am 12.09.2014).
- [19] André B. Bondi. „Characteristics of Scalability and Their Impact on Performance“. In: *Proceedings of the 2nd International Workshop on Software and Performance*. WOSP '00. Ottawa, Ontario, Kanada: ACM, 2000, S. 195–203.
- [20] Jürgen Bortz und Nicola Döring. *Forschungsmethoden und Evaluation*. 4. Auflage. Springer-Lehrbuch. Springer, 2006.
- [21] Jürgen Bortz und Christof Schuster. *Empirische Forschung und Skalenniveaus*. 7. Auflage. Springer-Lehrbuch. Springer, 2010.
- [22] Els Breugelmans, Clemens F. Köhler, Benedict G. C. Dellaert und Ko de Ruyter. „Promoting Interactive Decision Aids on Retail Websites: A Message Framing Perspective with New versus Traditional Focal Actions“. In: *Journal of Retailing* 88(2) (Juni 2012), S. 226–235.
- [23] Eric A. Brewer. *Towards robust distributed systems*. Principles of Distributed Computing (Invited Talk). Portland, Oregon, USA. Juli 2000.
- [24] Peter Buxmann, Thomas Hess und Sonja Lehmann. „Software as a Service“. In: *Wirtschaftsinformatik* 50(6) (2008), S. 500–503.
- [25] Giuseppe Carenini, Raymond T. Ng und Ed Zwart. „Extracting Knowledge from Evaluative Text“. In: *Proceedings of the 3rd International Conference on Knowledge Capture*. Banff, Alberta, Kanada: ACM, 2005, S. 11–18.
- [26] Andre Charland und Brian Leroux. „Mobile Application Development: Web vs. Native“. In: *Communications of the ACM* 54(5) (Mai 2011), S. 49–53.
- [27] Pei-Yu Chen, Shin-yi Wu und Jungsun Yoon. „The impact of online recommendations and consumer feedback on sales“. In: *Proceedings of the International Conference on Information Systems (ICIS)*. Washington, DC, USA, 2004, S. 711–724.
- [28] Xing Chen, Lin Li, Guandong Xu, Zhenglu Yang und Masaru Kitsuregawa. „Recommending Related Microblogs: A Comparison Between Topic and WordNet based Approaches“. In: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. Toronto, Kanada, 2012, S. 2417–2418.

- [29] Judith A. Chevalier und Dina Mayzlin. „The Effect of Word of Mouth on Sales: Online Book Reviews“. In: *Journal of Marketing Research* 43(3) (Aug. 2006), S. 345–354.
- [30] Eugene G. Chewning Jr. und Adrian M. Harrell. „The effect of information load on decision makers’ cue utilization levels and decision quality in a financial distress decision task“. In: *Accounting, Organizations and Society* 15(6) (1990), S. 527–542.
- [31] Frederick Chong und Gianpaolo Carraro. *Architecture Strategies for Catching the Long Tail*. Microsoft. Apr. 2006. URL: <http://msdn.microsoft.com/en-us/library/aa479069.aspx> (besucht am 29.04.2014).
- [32] Philipp Cimiano. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer-Verlag New York, 2006.
- [33] Jacob Cohen. „A power primer“. In: *Psychological bulletin* 112(1) (1992), S. 155–159.
- [34] William Cohen, Pradeep Ravikumar und Stephen Fienberg. „A comparison of string metrics for matching names and records“. In: *Proceedings of the Kdd workshop on data cleaning and object consolidation*. Bd. 3. American Association for Artificial Intelligence, 2003, S. 73–78.
- [35] Efthymios Constantinides und Stefan J. Fountain. „Web 2.0: Conceptual foundations and marketing issues“. In: *Journal of Direct, Data and Digital Marketing Practice* 9(3) (2008), S. 231–244.
- [36] Corinna Cortes und Vladimir Vapnik. „Support-vector networks“. In: *Machine Learning* 20(3) (1995), S. 273–297.
- [37] Renata Gonçalves Curty und Ping Zhang. „Social commerce: Looking back and forward“. In: *Proceedings of the American Society for Information Science and Technology* 48(1) (2011), S. 1–10.
- [38] Renata Gonçalves Curty und Ping Zhang. „Website features that gave rise to social commerce: a historical analysis“. In: *Electronic Commerce Research and Applications* 12(4) (2013), S. 260–279.
- [39] Fred D. Davis. „A technology acceptance model for empirically testing new end-user information systems – theory and results“. Diss. Massachusetts Inst. of Technology, 1985.
- [40] Klaas Dellschaft und Steffen Staab. „An Epistemic Dynamic Model for Tagging Systems“. In: *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*. Pittsburgh, PA, USA: ACM, 2008, S. 71–80.
- [41] Ravi Dhar. „The Effect of Decision Strategy on Deciding to Defer Choice“. In: *Journal of Behavioral Decision Making* 9(4) (1996), S. 265–281.
- [42] Xiaowen Ding, Bing Liu und Philip S. Yu. „A Holistic Lexicon-based Approach to Opinion Mining“. In: *Proceedings of the 2008 International Conference on Web Search and Data Mining*. WSDM ’08. Palo Alto, Kalifornien, USA: ACM, 2008, S. 231–240.
- [43] Pedro Domingos und Michael J. Pazzani. „Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier“. In: *Proceedings of the Thirteenth International Conference on Machine Learning*. Bari, Italien, Juli 1996, S. 105–112.

- [44] Wenjing Duan, Bin Gu und Andrew B. Whinston. „Do online reviews matter? — An empirical investigation of panel data“. In: *Decision Support Systems* 45(4) (Nov. 2008), S. 1007–1016.
- [45] Anja Ebersbach, Markus Glaser und Richard Heigl. *Social Web*. Konstanz: UVK Verlagsgesellschaft, 2008.
- [46] Werner Esswein, Sabine Zumpke und Nicole Sunke. „Identifying the Quality of e-Commerce Reference Models“. In: *Proceedings of the 6th International Conference on Electronic Commerce*. ICEC '04. Delft, Niederlande: ACM, 2004, S. 288–295.
- [47] Usama Fayyad, Gregory Piatetsky-Shapiro und Padhraic Smyth. „From data mining to knowledge discovery in databases“. In: *AI magazine* 17(3) (1996), S. 37–54.
- [48] Andy Field. *Discovering Statistics using IBM SPSS Statistics*. 4. Auflage. SAGE Publications, 2013.
- [49] Christiane Floyd. „A Systematic Look at Prototyping“. In: *Approaches to Prototyping*. Hrsg. von Reinhard Budde, Karin Kuhlenkamp, Lars Mathiassen und Heinz Züllighoven. Springer Berlin Heidelberg, 1984, S. 1–18.
- [50] Martin Fowler. *Patterns of Enterprise Application Architecture*. Addison-Wesley, 2003.
- [51] Fatih Gedikli und Dietmar Jannach. „Rating items by rating tags“. In: *Proceedings of the 2010 Workshop on Recommender Systems and the Social Web at ACM RecSys*. Barcelona, Spanien, 2010, S. 25–32.
- [52] Dave Gehrke und Efraim Turban. „Determinants of successful Website design: relative importance and recommendations for effectiveness“. In: *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*. Wailea Maui, Hawaii, USA, Jan. 1999, S. 1–8.
- [53] Scott A. Golder und Bernardo A. Huberman. „Usage patterns of collaborative tagging systems“. In: *Journal of information science* 32(2) (2006), S. 198–208.
- [54] Thomas R. Gruber. „A translation approach to portable ontology specifications“. In: *Knowledge Acquisition* 5(2) (1993), S. 199–220.
- [55] Chang Jie Guo, Wei Sun, Ying Huang, Zhi Hu Wang und Bo Gao. „A framework for native multi-tenancy application development and management“. In: *Proceedings of the 9th IEEE International Conference on E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*. 2007, S. 551–558.
- [56] Honglei Guo, Huijia Zhu, Zhili Guo, XiaoXun Zhang und Zhong Su. „Product Feature Categorization with Multilevel Latent Semantic Association“. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. Hong Kong, China: ACM, 2009, S. 1087–1096.
- [57] Manish Gupta, Rui Li, Zhijun Yin und Jiawei Han. „Survey on Social Tagging Techniques“. In: *ACM SIGKDD Explorations Newsletter* 12(1) (Juni 2010), S. 58–72.
- [58] Pranjal Gupta, Manjit S. Yadav und Rajan Varadarajan. „How Task-Facilitative Interactive Tools Foster Buyers’ Trust in Online Retailers: A Process View of

- Trust Development in the Electronic Marketplace“. In: *Journal of Retailing* 85(2) (2009), S. 159–176.
- [59] Harry Halpin, Valentin Robu und Hana Shepherd. „The complex dynamics of collaborative tagging“. In: *Proceedings of the 16th international conference on World Wide Web*. Banff, Alberta, Kanada: ACM, 2007, S. 211–220.
- [60] Zellig Harris. *Mathematical structures of language*. John Wiley & Sons, 1968.
- [61] Trevor Hastie und Robert Tibshirani. „Classification by pairwise coupling“. In: *The annals of statistics* 26(2) (1998), S. 451–471.
- [62] Gerald Häubl und Valerie Trifts. „Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids“. In: *Marketing Science* 19(1) (Jan. 2000), S. 4–21.
- [63] Haufe. *Das erwarten Kunden von Webshops*. URL: http://www.haufe.de/marketing-vertrieb/vertrieb/e-commerce-das-erwarten-kunden-von-webshops_130_213778.html (besucht am 20.03.2014).
- [64] Marti Hearst. „Design recommendations for hierarchical faceted search interfaces“. In: *Proceedings of the ACM SIGIR workshop on faceted search*. Seattle, WA, USA, 2006, S. 1–5.
- [65] Markus Heckner, Michael Heilemann und Christian Wolff. „Personal Information Management vs. Resource Sharing: Towards a Model of Information Behavior in Social Tagging Systems“. In: *Proceedings of the Third International ICWSM Conference*. San Jose, CA, USA, 2009.
- [66] Gerrit Heinemann. *Der neue Online-Handel – Geschäftsmodell und Kanalexzellenz im E-Commerce*. Springer Fachmedien Wiesbaden, 2014.
- [67] Martin Hepp, Katharina Siorpaes und Daniel Bachlechner. „Harvesting wiki consensus: Using wikipedia entries as vocabulary for knowledge management“. In: *Internet Computing, IEEE* 11(5) (2007), S. 54–65.
- [68] Serena Hillman, Carman Neustaedter, John Bowes und Alissa Antle. „Soft Trust and mCommerce Shopping Behaviours“. In: *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*. San Francisco, Kalifornien, USA: ACM, 2012, S. 113–122.
- [69] Hajo Hippner. „Bedeutung, Anwendungen und Einsatzpotenziale von Social Software“. In: *HMD – Praxis der Wirtschaftsinformatik* 43(252) (2006), S. 6–16.
- [70] Stephen J. Hoch, Eric T. Bradlow und Brian Wansink. „The Variety of An Assortment“. In: *Marketing Science* 18(4) (Apr. 1999), S. 527–546.
- [71] Clyde Holsapple. „DSS Architecture and Types“. In: *Handbook on Decision Support Systems 1*. International Handbooks Information System. Springer Berlin Heidelberg, 2008, S. 163–189.
- [72] Andreas Hotho, Steffen Staab und Gerd Stumme. „Wordnet improves Text Document Clustering“. In: *Proceedings of the Semantic Web Workshop of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval (SIGIR 2003)*. Hrsg. von Ying Ding, Keith van Rijsbergen, Iadh Ounis und Joemon Jose. Toronto, Kanada, 2003.

- [73] Minqing Hu und Bing Liu. „Mining and summarizing customer reviews“. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. Seattle, WA, USA, 2004, S. 168–177.
- [74] Minqing Hu und Bing Liu. „Mining Opinion Features in Customer Reviews“. In: *Proceedings of the 19th National Conference on Artificial Intelligence*. San Jose, Kalifornien: AAAI Press, 2004, S. 755–760.
- [75] Minqing Hu und Bing Liu. „Opinion Feature Extraction Using Class Sequential Rules“. In: *Proceedings of the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. 2006, S. 61–66.
- [76] Nan Hu, Indranil Bose, Noi Sian Koh und Ling Liu. „Manipulation of online reviews: An analysis of ratings, readability, and sentiments“. In: *Decision Support Systems* 52(3) (2012), S. 674–684.
- [77] Nan Hu, Noi Sian Koh und Srinivas K. Reddy. „Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales“. In: *Decision Support Systems* 57 (2014), S. 42–53.
- [78] Xia Hu, Nan Sun, Chao Zhang und Tat-Seng Chua. „Exploiting Internal and External Semantics for the Clustering of Short Texts Using World Knowledge“. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. Hong Kong, China, 2009, S. 919–928.
- [79] Jeff Huang, Katherine M. Thornton und Efthimis N. Efthimiadis. „Conversational Tagging in Twitter“. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. HT ’10. Toronto, Ontario, Kanada: ACM, 2010, S. 173–178.
- [80] Zhao Huang und Morad Benyoucef. „From e-commerce to social commerce: A close look at design features“. In: *Electronic Commerce Research and Applications* 12(4) (2013), S. 246–259.
- [81] Zhao Huang, Seo Yeon Yoon und Morad Benyoucef. „Adding Social Features to E-commerce“. In: *Proceedings of the Conference on Information Systems Applied Research*. New Orleans, Louisiana, USA, 2012, S. 1–11.
- [82] Ross Ihaka und Robert Gentleman. „R: a language for data analysis and graphics“. In: *Journal of computational and graphical statistics* 5(3) (1996), S. 299–314.
- [83] Aminul Islam und Diana Inkpen. „Semantic Text Similarity Using Corpus-based Word Similarity and String Similarity“. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 2(2) (Juli 2008), 10:1–10:25.
- [84] Susan Jamieson. „Likert scales: how to (ab)use them“. In: *Medical education* 38(12) (2004), S. 1217–1218.
- [85] Lee Jensen und Tony Martinez. „Improving Text Classification by Using Conceptual and Contextual Features“. In: *Proceedings of the Workshop on Text Mining at the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2000, S. 101–102.
- [86] Thorsten Joachims. „Text categorization with Support Vector Machines: Learning with many relevant features“. In: *Machine Learning: ECML-98*. Hrsg. von Claire Nédellec und Céline Rouveirol. Bd. 1398. Lecture Notes in Computer Science. Springer, 1998, S. 137–142.

- [87] Eric J. Johnson und John W. Payne. „Effort and Accuracy in Choice“. In: *Management Science* 31(4) (1985), S. 395–414.
- [88] Daniel Kailer und Peter Mandl. „Combining Hashtags and Ratings for Opinion Expression in Customer Reviews“. In: *Proceedings of the 13th International Conference on WWW/Internet*. IADIS. Porto, Portugal, 2014, S. 352–356.
- [89] Daniel Kailer und Peter Mandl. „Hoch verfügbare und konsistente Datenhaltung in der Cloud“. In: *HMD – Praxis der Wirtschaftsinformatik* 49(6) (2012), S. 69–77.
- [90] Daniel Kailer, Peter Mandl und Alexander Schill. „An Empirical Study on the Usage of Social Media in German B2C-Online Stores“. In: *International journal of advanced Information technology* 3(5) (2013), S. 1–14.
- [91] Daniel Kailer, Peter Mandl und Alexander Schill. „Grouping Product Aspects from Short Texts Using Multiple Classifiers“. In: *Proceedings of the 16th International Conference on Web Information Systems Engineering (WISE)*. Springer Lecture Notes in Computer Science 9418. Miami, FL, USA, Nov. 2015, S. 1–15.
- [92] Daniel Kailer, Peter Mandl und Alexander Schill. „Rated Tags: Adding rating capability to collaborative tagging“. In: *Proceedings of the Third International Conference on Social Computing and Its Applications*. IEEE. Karlsruhe, 2013, S. 249–255.
- [93] Daniel Kailer, Peter Mandl und Alexander Schill. „Supporting customers’ decision making with Rated Tags“. In: *Proceedings of the 16th International Conference on Electronic Commerce*. ACM. Philadelphia, PA, USA, 2014, S. 33–40.
- [94] Arnold Kamis und Michael J. Davern. „Personalizing to product category knowledge: exploring the mediating effect of shopping tools on decision confidence“. In: *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. Waikoloa, Hawaii, USA, 2004, S. 1–10.
- [95] Anne Kao und Stephen R. Poteet. „Overview“. In: *Natural Language Processing and Text Mining*. Hrsg. von Anne Kao und Stephen R. Poteet. Springer London, 2007, S. 1–7.
- [96] Andreas M. Kaplan und Michael Haenlein. „Users of the world, unite! The challenges and opportunities of Social Media“. In: *Business Horizons* 53(1) (Jan. 2010), S. 59–68.
- [97] Jan H. Kietzmann, Kristopher Hermkens, Ian P. McCarthy und Bruno S. Silvestre. „Social media? Get serious! Understanding the functional building blocks of social media“. In: *Business Horizons* 54(3) (2011), S. 241–251.
- [98] Young Ae Kim und Jaideep Srivastava. „Impact of Social Influence in e-Commerce Decision Making“. In: *Proceedings of the Ninth International Conference on Electronic Commerce*. ICEC ’07. Minneapolis, MN, USA: ACM, 2007, S. 293–302.
- [99] Josef Kittler, Mohamad Hatef, Robert P.W. Duin und Jiri Matas. „On combining classifiers“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(3) (1998), S. 226–239.
- [100] André Klahold. *Empfehlungssysteme*. Wiesbaden: Vieweg + Teubner, 2009.
- [101] Jonathan Koren, Yi Zhang und Xue Liu. „Personalized interactive faceted search“. In: *Proceedings of the 17th international conference on World Wide Web*. ACM. 2008, S. 477–486.

- [102] Christian Körner, Roman Kern, Hans-Peter Grahsl und Markus Strohmaier. „Of Categorizers and Describers: An Evaluation of Quantitative Measures for Tagging Motivation“. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. HT '10. Toronto, Ontario, Kanada: ACM, 2010, S. 157–166.
- [103] Philip Kotler und Friedhelm Bliemel. *Marketing-Management*. 12. Auflage. Pearson Verlag, 2007.
- [104] Philip Kotler und Kevin Lane Keller. *Marketing Management*. 14. Auflage. Prentice Hall, 2012.
- [105] Rouven Krebs, Christof Momm und Samuel Kounev. „Architectural Concerns in Multi-tenant SaaS Applications“. In: *Proceedings of the 2nd International Conference on Cloud Computing and Service Science*. 2012, S. 426–431.
- [106] Philippe B. Kruchten. „The 4+1 View Model of architecture“. In: *IEEE Software* 12(6) (Nov. 1995), S. 42–50.
- [107] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [108] Thomas Kwok, Thao Nguyen und Linh Lam. „A Software as a Service with Multi-tenancy Support for an Electronic Contract Management Application“. In: *Proceedings of the IEEE International Conference on Services Computing*. Bd. 2. Honolulu, HI, USA, Juli 2008, S. 179–186.
- [109] Georg Lackermair. „Hybrid cloud architectures for the online commerce“. In: *Procedia Computer Science* 3 (2011). World Conference on Information Technology, S. 550–555.
- [110] Georg Lackermair, Daniel Kailer und Kenan Kanmaz. „Importance of Online Product Reviews from a Customer’s Perspective“. In: *Advances in Economics and Business* 1(1) (2013), S. 1–5.
- [111] Georg Lackermair und Jakob Reuder. „Nutzung interaktiver Elemente in deutschen Online-Shops“. In: *Communities in New Media: Virtual Enterprises, Research Communities & Social Media Networks*. Dresden: TUDpress, 2012, S. 207–214.
- [112] Erwin Lammenett. *Praxiswissen Online-Marketing*. 4. Auflage. Springer Fachmedien Wiesbaden, 2014.
- [113] Kenneth C. Laudon, Jane P. Laudon und Detlef Schoder. *Wirtschaftsinformatik – Eine Einführung*. 2. Auflage. Pearson Studium. Pearson Deutschland, 2010.
- [114] Kenneth C. Laudon und Carol Guercio Traver. *E-Commerce 2013 – business. technology. society*. 9. Auflage. Prentice Hall, 2013.
- [115] Helmut Laux und Felix Liermann. *Grundlagen der Organisation – Die Steuerung von Entscheidungen als Grundproblem der Betriebswirtschaftslehre*. Springer Berlin Heidelberg, 2005.
- [116] Sung Eob Lee, Dong Kwan Son und Steve SangKi Han. „Qtag: Tagging as a Means of Rating, Opinion-Expressing, Sharing and Visualizing“. In: *Proceedings of the 25th Annual ACM International Conference on Design of Communication*. El Paso, Texas, USA, 2007, S. 189–195.
- [117] Peter Leitner und Thomas Grechenig. *Social Online Shopping: Neue Formen der Interaktion und Kollaboration im Electronic Commerce der Zukunft*. In: *Wirtschaftsinformatik Proceedings 2009*, Paper 105, S. 243–252. 2009.

- [118] Bing Liu. *Web Data Mining. Data-Centric Systems and Applications*. Springer Berlin Heidelberg, 2007.
- [119] Bing Liu, Mingqing Hu und Junsheng Cheng. „Opinion Observer: Analyzing and Comparing Opinions on the Web“. In: *Proceedings of the 14th International Conference on World Wide Web. WWW '05*. Chiba, Japan: ACM, 2005, S. 342–351.
- [120] Bing Liu und Lei Zhang. „A Survey of Opinion Mining and Sentiment Analysis“. In: *Mining Text Data*. Hrsg. von Charu C. Aggarwal und ChengXiang Zhai. Springer US, 2012, S. 415–463.
- [121] Qianqian (Ben) Liu, Elena Karahanna und Richard T. Watson. „Unveiling user-generated content: Designing websites to best present customer reviews“. In: *Business Horizons* 54(3) (Mai 2011), S. 231–240.
- [122] Ying Liu, Han Tong Loh, Kamal Youcef-Toumi und Shu Beng Tor. „Handling of Imbalanced Data in Text Classification: Category-Based Term Weights“. In: *Natural Language Processing and Text Mining*. Hrsg. von Anne Kao und Stephen R. Poteet. Springer London, 2007, S. 171–192.
- [123] Gerald L. Lohse und Eric J. Johnson. „A comparison of two process tracing methods for choice tasks“. In: *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences*. Bd. 4. IEEE. 1996, S. 86–97.
- [124] Anália G. Lourenço und Orlando O. Belo. „Catching Web Crawlers in the Act“. In: *Proceedings of the 6th International Conference on Web Engineering. ICWE '06*. Palo Alto, Kalifornien, USA: ACM, 2006, S. 265–272.
- [125] Yue Lu, ChengXiang Zhai und Neel Sundaresan. „Rated Aspect Summarization of Short Comments“. In: *Proceedings of the 18th International Conference on World Wide Web*. Madrid, Spanien: ACM, 2009, S. 131–140.
- [126] C. Matthew MacKenzie, Ken Laskey, Francis McCabe, Peter F. Brown und Rebekah Metz. *Reference model for service oriented architecture 1.0*. OASIS. Aug. 2006.
- [127] Christopher D. Manning, Prabhakar Raghavan und Hinrich Schütze. *Introduction to Information Retrieval*. Online Edition. Cambridge University Press, 2009.
- [128] Trevor Mansuy und Robert J. Hilderman. „A Characterization of Wordnet Features in Boolean Models for Text Classification“. In: *Proceedings of the Fifth Australasian Conference on Data Mining and Analytics - Volume 61. AusDM '06*. Sydney, Australien: Australian Computer Society, 2006, S. 103–109.
- [129] Cameron Marlow, Mor Naaman, Danah Boyd und Marc Davis. „HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read“. In: *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia. HYPERTEXT '06*. Odense, Dänemark: ACM, 2006, S. 31–40.
- [130] Paul Marsden. „eBranding and Social Commerce“. In: *Brand Evolution*. Hrsg. von Elke Theobald und Philipp T. Haisch. Gabler, 2011, S. 357–372.
- [131] Andrew McCallum und Kamal Nigam. „A comparison of event models for naive bayes text classification“. In: *Proceedings of the AAAI-98 workshop on learning for text categorization*. AAAI Press, 1998, S. 41–48.

- [132] Donald Metzler, Susan Dumais und Christopher Meek. „Similarity Measures for Short Segments of Text“. In: *Proceedings of the 29th European Conference on IR Research*. ECIR'07. Rom, Italien: Springer-Verlag, 2007, S. 16–27.
- [133] Maged Michael, Jose E. Moreira, Doron Shiloach und Robert W. Wisniewski. „Scale-up x scale-out: A case study using nutch/lucene“. In: *Proceedings of the Parallel and Distributed Processing Symposium*. Long Beach, CA, USA: IEEE, 2007, S. 1–8.
- [134] Rada Mihalcea, Courtney Corley und Carlo Strapparava. „Corpus-based and Knowledge-based Measures of Text Semantic Similarity“. In: *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*. Boston, Massachusetts, USA: AAAI Press, 2006, S. 775–780.
- [135] George A. Miller. „WordNet: A Lexical Database for English“. In: *Communications of the ACM* 38(11) (Nov. 1995), S. 39–41.
- [136] Vincent-Wayne Mitchell, Gianfranco Walsh und Mo Yami. „Towards a Conceptual Model of Consumer Confusion“. In: *Advances in Consumer Research* 32(1) (2005), S. 143–150.
- [137] Sanjay Mohapatra. *E-Commerce Strategy*. New York: Springer, 2013.
- [138] Kyle B. Murray und Gerald Häubl. „Interactive Consumer Decision Aids“. In: *Handbook of Marketing Decision Models*. Hrsg. von Berend Wierenga. Bd. 121. International Series in Operations Research & Management Science. Springer US, 2008, S. 55–77.
- [139] Phillip Nelson. „Advertising as Information“. In: *The Journal of Political Economy* 83 (1974), S. 729–754.
- [140] Jürgen Noll. „Marketingfaktoren, Kaufentscheidungsprozess und Kundenbindung im e-commerce“. In: *der markt* 42(164) (2003), S. 35–44.
- [141] Oded Nov, Mor Naaman und Chen Ye. „What Drives Content Tagging: The Case of Photos on Flickr“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. Florenz, Italien: ACM, 2008, S. 1097–1100.
- [142] Robert M. O’Keefe und Tim McEachern. „Web-based Customer Decision Support Systems“. In: *Communications of the ACM* 41(3) (März 1998), S. 71–78.
- [143] David Opitz und Richard Maclin. „Popular ensemble methods: An empirical study“. In: *Journal of Artificial Intelligence Research* 11 (1999), S. 169–198.
- [144] Tim O’Reilly. *Web 2.0 Compact Definition: Trying Again*. Dez. 2006. URL: <http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html> (besucht am 05.03.2014).
- [145] Tim O’Reilly. *What Is Web 2.0*. Sep. 2005. URL: <http://oreilly.com/web2/archive/what-is-web-20.html> (besucht am 05.03.2014).
- [146] Muzaffer Ozakca und Youn-Kyung Lim. „A Study of Reviews and Ratings on the Internet“. In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '06. Montréal, Québec, Kanada: ACM, 2006, S. 1181–1186.
- [147] Bo Pang und Lillian Lee. „Opinion Mining and Sentiment Analysis“. In: *Foundations and Trends in Information Retrieval* 2(1–2) (2008), S. 1–135.
- [148] Bo Pang und Lillian Lee. „Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales“. In: *Proceedings of the 43rd Annual*

- Meeting on Association for Computational Linguistics*. Ann Arbor, Michigan, USA, 2005, S. 115–124.
- [149] Michael P. Papazoglou, Paolo Traverso, Schahram Dustdar und Frank Leymann. „Service-oriented computing: a research roadmap“. In: *International Journal of Cooperative Information Systems* 17(2) (2008), S. 223–255.
 - [150] Do-Hyung Park und Jumin Lee. „eWOM overload and its effect on consumer behavioral intention depending on consumer involvement“. In: *Electronic Commerce Research and Applications* 7(4) (Dez. 2008), S. 386–398.
 - [151] Do-Hyung Park, Jumin Lee und Ingoo Han. „The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement“. In: *International Journal of Electronic Commerce* 11(4) (2007), S. 125–148.
 - [152] John W. Payne. „Contingent decision behavior“. In: *Psychological Bulletin* 92(2) (1982), S. 382–402.
 - [153] John W. Payne, James R. Bettman, Eloise Coupey und Eric J. Johnson. „A constructive process view of decision making: Multiple strategies in judgment and choice“. In: *Acta Psychologica* 80(1–3) (1992), S. 107–141.
 - [154] John W. Payne, James R. Bettman und Eric J. Johnson. *The Adaptive Decision Maker*. Cambridge University Press, 1993.
 - [155] Jella Pfeiffer. *Interactive Decision Aids in E-Commerce*. Heidelberg: Springer, 2012.
 - [156] Lars Pind. *Folksonomies: How we can improve the tags*. Jan. 2005. URL: <http://calvinconaway.com/2005/01/23/folksonomies-how-we-can-improve-the-tags> (besucht am 05.04.2014).
 - [157] Robi Polikar. „Ensemble based systems in decision making“. In: *Circuits and systems magazine* 6(3) (2006), S. 21–45.
 - [158] Gustav Pomberger und Wolfgang Pree. *Software Engineering: Architektur-Design und Prozessorientierung*. 3. Auflage. Hanser, 2004.
 - [159] Ana-Maria Popescu und Orena Etzioni. „Extracting Product Features and Opinions from Reviews“. In: *Natural Language Processing and Text Mining*. Hrsg. von Anne Kao und Stephen R. Poteet. Springer London, 2007, S. 9–28.
 - [160] Martin F. Porter. „An algorithm for suffix stripping“. In: *Program* 14(3) (1980), S. 130–137.
 - [161] Daniel J. Power und Shashidhar Kaparathi. „Building Web-based decision support systems“. In: *Studies in Informatics and Control* 11(4) (2002), S. 291–302.
 - [162] Ayesha Rashid, Naveed Anwer, Muddaser Iqbal und Muhammad Sher. „A Survey Paper: Areas, Techniques and Challenges of Opinion Mining“. In: *International Journal of Computer Science Issues* 10(2) (2013), S. 18–31.
 - [163] Jason D. M. Rennie, Lawrence Shih, Jaime Teevan und David R. Karger. „Tackling the poor assumptions of naive bayes text classifiers“. In: *Proceedings of the Twentieth International Conference on Machine Learning*. Bd. 3. Washington DC, USA, 2003, S. 616–623.
 - [164] Francesco Ricci, Lior Rokach und Bracha Shapira. „Introduction to Recommender Systems Handbook“. In: *Recommender Systems Handbook*. Hrsg. von Francesco

- Ricci, Lior Rokach, Bracha Shapira und Paul B. Kantor. Springer US, 2011, S. 1–35.
- [165] Francesco Ricci und René T. A. Wietsma. „Product Reviews in Travel Decision Making“. In: *Information and Communication Technologies in Tourism 2006*. Hrsg. von Martin Hitz, Marianna Sigala und Jamie Murphy. Springer Vienna, 2006, S. 296–307.
- [166] Alexander Richter und Michael Koch. *Social Software – Status quo und Zukunft*. Techn. Ber. 2007-01. Universität der Bundeswehr München, Fakultät für Informatik, Feb. 2007.
- [167] René Riedl, Eduard Brandstätter und Friedrich Roithmayr. „Identifying decision strategies: A process-and outcome-based classification method“. In: *Behavior Research Methods* 40(3) (2008), S. 795–807.
- [168] John Ben Schafer, Joseph Konstan und John Riedl. „Recommender Systems in e-Commerce“. In: *Proceedings of the 1st ACM Conference on Electronic Commerce*. EC ’99. Denver, Colorado, USA: ACM, 1999, S. 158–166.
- [169] Alexander Schill und Thomas Springer. *Verteilte Systeme. Grundlagen und Basistechnologien*. 2. Auflage. eXamen.press. Springer Berlin Heidelberg, 2012.
- [170] Jan-Hinrik Schmidt. *Social Media*. Springer Fachmedien Wiesbaden, 2013.
- [171] Michael Scholz und Verena Dorner. „Das Rezept für die perfekte Rezension?“. In: *WIRTSCHAFTSINFORMATIK* 55(3) (2013), S. 135–146.
- [172] Sam Scott und Stan Matwin. „Text classification using WordNet hypernyms“. In: *Proceedings of the Workshop on Usage of WordNet in Natural Language Processing Systems*. Montreal, Kanada, 1998, S. 38–44.
- [173] Fabrizio Sebastiani. „Machine Learning in Automated Text Categorization“. In: *ACM Computing Surveys* 34(1) (März 2002), S. 1–47.
- [174] Bakhtawar Seerat und Farouque Azam. „Opinion Mining: Issues and Challenges (A survey)“. In: *International Journal of Computer Applications* 49(9) (Juli 2012), S. 42–51.
- [175] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper und John Riedl. „Tagging, Communities, Vocabulary, Evolution“. In: *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work*. CSCW ’06. Banff, Alberta, Kanada: ACM, 2006, S. 181–190.
- [176] Michelle M. H. Şeref und Ravindra K. Ahuja. „Spreadsheet-Based Decision Support Systems“. In: *Handbook on Decision Support Systems 1*. International Handbooks Information System. Springer Berlin Heidelberg, 2008, S. 277–298.
- [177] Jung P. Shim, Merrill Warkentin, James F. Courtney, Daniel J. Power, Ramesh Sharda und Christer Carlsson. „Past, present, and future of decision support technology“. In: *Decision Support Systems* 33(2) (2002), S. 111–126.
- [178] Ben Shneiderman und Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 5. Auflage (international edition). Pearson Addison Wesley, 2010.
- [179] Steven M. Shugan. „The Cost of Thinking“. In: *The Journal of Consumer Research* 7(2) (1980), S. 99–111.

- [180] Gene Smith. *Tagging: people-powered metadata for the social web*. Berkeley, CA: New Riders, 2008.
- [181] Ian Sommerville. *Software Engineering*. 9. Auflage (international edition). Harlow, England: Pearson, 2010.
- [182] Sarah Spiekermann und Corina Paraschiv. „Motivating Human-Agent Interaction: Transferring Insights from Behavioral Marketing to Interface Design“. In: *Electronic Commerce Research* 2(3) (2002), S. 255–285.
- [183] Gernot Starke. *Effektive Software-Architekturen*. Hanser, 2011.
- [184] Statista. *Anteil des E-Commerce am Einzelhandelsumsatz in Deutschland von 2009 bis 2014*. März 2015. URL: <http://de.statista.com/statistik/daten/studie/201859/umfrage/anteil-des-e-commerce-am-einzelhandelsumsatz> (besucht am 24.06.2015).
- [185] Jonathan Steuer. „Defining Virtual Reality: Dimensions Determining Telepresence“. In: *Journal of Communication* 42(4) (1992), S. 73–93.
- [186] Henrik Stormer und Daniel Frauchinger. „Aktuelle Entwicklungen elektronischer Shopsysteme“. In: *HMD - Praxis der Wirtschaftsinformatik* 45(261) (2008), S. 61–70.
- [187] David Streitfeld. *Giving Mom’s Book Five Stars? Amazon May Cull Your Review*. Dez. 2012. URL: <http://www.nytimes.com/2012/12/23/technology/amazon-book-reviews-deleted-in-a-purge-aimed-at-manipulation.html> (besucht am 05.04.2014).
- [188] Markus Strohmaier, Christian Körner und Roman Kern. „Understanding why users tag: A survey of tagging motivation literature and results from an empirical study“. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 17 (2012), S. 1–11.
- [189] Qi Su, Xinying Xu, Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang, Bin Swen und Zhong Su. „Hidden Sentiment Association in Chinese Web Opinion Mining“. In: *Proceedings of the 17th International Conference on World Wide Web*. Beijing, China: ACM, 2008, S. 959–968.
- [190] Clemens Szyperski. *Component Software: Beyond Object-Oriented Programming*. 2. Auflage. Boston, MA, USA: Addison-Wesley, 2002.
- [191] Wee-Kek Tan, Chuan-Hoo Tan und Hock-Hai Teo. „Consumer-based decision aid that explains which to buy: Decision confirmation or overconfidence bias?“ In: *Decision Support Systems* 53(1) (Apr. 2012), S. 127–141.
- [192] Peter Todd und Izak Benbasat. „Inducing Compensatory Information Processing through Decision Aids that Facilitate Effort Reduction: An Experimental Assessment“. In: *Journal of Behavioral Decision Making* 13(1) (2000), S. 91–106.
- [193] Matthias F. Treutner und Herwig Ostermann. „Evolution of Standard Web Shop Software Systems: A Review and Analysis of Literature and Market Surveys“. In: *The Open Information Systems Journal* 5 (2011), S. 8–18.
- [194] TripAdvisor. *TripAdvisor Reaches 100 Million Reviews And Opinions Milestone*. März 2013. URL: http://www.tripadvisor.com/PressCenter-i5799-c1-Press_Releases.html (besucht am 11.04.2014).

- [195] Volker Trommsdorff. *Konsumentenverhalten*. 7. Auflage. W. Kohlhammer Verlag, 2009.
- [196] Efraim Turban. *Decision Support and Expert Systems: Management Support Systems*. 2. Auflage. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1990.
- [197] Eifraim Turban, David King, Jae Lee, Ting-Peng Liang und Deborrah Turban. *Electronic Commerce 2010: A Managerial Perspective*. 6. Auflage. Pearson Education, 2010.
- [198] Peter D. Turney. „Mining the Web for Synonyms: PMI-IR Versus LSA on TOEFL“. In: *Proceedings of the 12th European Conference on Machine Learning*. London, UK: Springer-Verlag, 2001, S. 491–502.
- [199] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [200] Jesse Vig, Matthew Soukup, Shilad Sen und John Riedl. „Tag Expression: Tagging with Feeling“. In: *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*. UIST '10. New York, NY, USA: ACM, 2010, S. 323–332.
- [201] Leo R. Vijayarathy und Joseph M. Jones. „Do Internet Shopping Aids Make a Difference? An Empirical Investigation“. In: *Electronic Markets* 11(1) (2001), S. 75–83.
- [202] Werner Vogels. „Eventually Consistent“. In: *Communications of the ACM* 52(1) (Jan. 2009), S. 40–44.
- [203] Dirk Vorberg und Sven Blankenberger. „Die Auswahl statistischer Tests und Maße“. In: *Psychologische Rundschau* 50(3) (1999), S. 157–164.
- [204] Robert A. Wagner und Michael J. Fischer. „The String-to-String Correction Problem“. In: *Journal of the Association for Computing Machinery* 21(1) (Jan. 1974), S. 168–173.
- [205] Gertrud Walgenbach. *Die Vorteilssituation von Innovatoren auf elektronischen Märkten*. Gabler, 2007.
- [206] Chingning Wang und Ping Zhang. „The Evolution of Social Commerce: An Examination from the People, Business, Technology, and Information Perspective“. In: *Communications of the AIS* 31(1) (2012), S. 105–127.
- [207] Hongning Wang, Yue Lu und Chengxiang Zhai. „Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach“. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '10. Washington, DC, USA: ACM, 2010, S. 783–792.
- [208] Weiquan Wang und Izak Benbasat. „Interactive Decision Aids for Consumer Decision Making in E-commerce: The Influence of Perceived Strategy Restrictiveness“. In: *MIS Quarterly* 33(2) (Juni 2009), S. 293–320.
- [209] Zhi Hu Wang, Chang Jie Guo, Bo Gao, Wei Sun, Zhen Zhang und Wen Hao An. „A Study and Performance Evaluation of the Multi-Tenant Data Tier Design Patterns for Service Oriented Computing“. In: *Proceedings of the IEEE International Conference on e-Business Engineering*. Okt. 2008, S. 94–101.

- [210] Chris M. White und Ulrich Hoffrage. „Testing the tyranny of too much choice against the allure of more choice“. In: *Psychology and Marketing* 26(3) (2009), S. 280–298.
- [211] Klaus-Peter Wiedmann, Sascha Langner und Jens Friedlandt. „Welche Kundenrezensionen werden gelesen?“ In: *Das Internet der Zukunft*. Hrsg. von Udo Wagner, Klaus-Peter Wiedmann und Dietrich Oelsnitz. Gabler, 2011, S. 329–349.
- [212] Erik Wiener, Jan O. Pedersen und Andreas S. Weigend. „A neural network approach to topic spotting“. In: *Proceedings of SDAIR-95, 4th annual symposium on document analysis and information retrieval*. Las Vegas, Nevada, USA, 1995, S. 317–332.
- [213] Rolf T. Wigand, Robert I. Benjamin und Johanna L. H. Birkland. „Web 2.0 and Beyond: Implications for Electronic Commerce“. In: *Proceedings of the 10th International Conference on Electronic Commerce*. ICEC '08. Innsbruck, Österreich: ACM, 2008, 7:1–7:5.
- [214] Bo Xiao und Izak Benbasat. „E-Commerce product recommendation agents: use, characteristics, and impact“. In: *MIS Quarterly* 31(1) (2007), S. 137–209.
- [215] Lei Xu, Adam Krzyzak und Ching Y. Suen. „Methods of combining multiple classifiers and their applications to handwriting recognition“. In: *IEEE Transactions on Systems, Man and Cybernetics* 22(3) (Mai 1992), S. 418–435.
- [216] Zhichen Xu, Yun Fu, Jianchang Mao und Difu Su. „Towards the semantic web: Collaborative tag suggestions“. In: *Proceedings of Collaborative Web Tagging Workshop at 15th International World Wide Web Conference*. Edinburgh, Schottland, 2006.
- [217] Nick Yeung und Christopher Summerfield. „Metacognition in human decision-making: confidence and error monitoring“. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367(1594) (2012), S. 1310–1321.
- [218] Fatemeh Zahedi, Jaeki Song und Suprasith Jarupathirun. „Web-Based Decision Support“. In: *Handbook on Decision Support Systems 1*. International Handbooks Information System. Springer Berlin Heidelberg, 2008, S. 315–338.
- [219] Zhongwu Zhai, Bing Liu, Hua Xu und Peifa Jia. „Grouping product features using semi-supervised learning with soft-constraints“. In: *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics. Beijing, China, Aug. 2010, S. 1272–1280.
- [220] Lina Zhou, Ping Zhang und Hans-Dieter Zimmermann. „Social commerce research: An integrated view“. In: *Electronic Commerce Research and Applications* 12(2) (2013), S. 61–68.

Abbildungsverzeichnis

1.1	Anteil des E-Commerce am Einzelhandelsumsatz in Deutschland	10
1.2	Übersicht der Kernkapitel dieser Arbeit	13
2.1	Beispielhafte Architektur eines Onlineshop-Systems auf hoher Ebene . . .	17
2.2	Phasen eines Entscheidungsprozesses im E-Commerce	19
2.3	Abnahme der Entscheidungsleistung ab einer gewissen Menge an Informa- tionen	20
2.4	Komponenten eines DSS	22
2.5	Kategorisierung von Entscheidungshilfen	23
2.6	Beispielhafte Vergleichstabelle zweier Monitore	26
2.7	Einordnung des Themas im Bereich des E-Commerce	27
2.8	Einordnung des Themas im Bereich Entscheidungsunterstützung	27
3.1	Bausteine von Social Media	31
3.2	Sortimentsverteilung der untersuchten Onlineshops	36
3.3	Histogramm und Boxplot für die Verteilung der Social Media Features in den untersuchten Onlineshops	38
3.4	Verteilung der gefundenen Social Media Features auf die sieben Bausteine des Social Media	42
3.5	Detaillierte Bewertung von vordefinierten Produktaspekten	44
3.6	Beispielhaftes, vereinfachtes Modell eines Social Tagging Systems	48
3.7	Einordnung des Themas im Bereich Social Media	52
4.1	Beispiel für die geringe Aussagekraft einer Gesamtbewertung	54
4.2	Einordnung des Rated Tags-Modells in verwandte Themenbereiche	55
4.3	Beispielhafte Darstellung des Rated Tags-Ansatzes	57
4.4	Beispielszenario zur Verwendung von Rated Tags, um die Entscheidungs- findung zu verbessern	58
4.5	UML-Anwendungsfalldiagramm für das Rated Tags-Modell	59
4.6	UML-Aktivitätsdiagramm zur beispielhaften Darstellung eines Entschei- dungsprozesses	62
4.7	Vereinfachte Darstellung der relevanten Entitäten und Beziehungen des Rated Tags-Modells in Form eines ER-Diagramms	64
4.8	Beispiel einer Gesamtbewertung inklusive der Angabe von Vor- und Nach- teilen	69

5.1	Einordnung der Rated Tags-Komponenten in die beispielhafte Systemarchitektur eines Onlineshop-Systems aus Abbildung 2.1	76
5.2	Drei-Schichten-Architektur des Rated Tags-Systems	78
5.3	Abstrakte Darstellung der Client-seitigen Pakete des Rated Tags-Systems basierend auf dem MVC-Entwurfsmuster	79
5.4	Abstrakte Darstellung der Server-seitigen Pakete des Rated Tags-Systems	80
5.5	UML-Sequenzdiagramm für die Erstellung eines Reviews	84
5.6	Beispielhaftes UML-Sequenzdiagramm für eine Filterung von Rezensionen zur Laufzeit	85
5.7	Beispielhafte Facettensuche auf Amazon.de	88
5.8	Technologiestack und logisches 3-Schichtenarchitektur des Prototyps	90
5.9	Beispielhafter Auszug der traditionellen Elemente einer Rezension	93
5.10	Beispielhafte Zuweisung von Rated Tags an eine Rezension	94
5.11	Beispielhafte Darstellung der in Abbildung 5.9 erzeugten Rezension	94
5.12	Filtermöglichkeit von Rezensionen anhand von Rated Tags	94
6.1	Forschungsmodell der Anwenderstudie zur Evaluierung des Rated Tags-Ansatzes	96
6.2	Beispielhafter Ausschnitt einer Produktansicht für Teilnehmer der Rated Tags-Gruppe	97
6.3	Beispielhafte Rezension sowie die dafür erstellten Rated Tags	98
6.4	Einordnung der Studie zur Bestimmung des statistischen Testverfahrens	102
6.5	Verteilung der Altersklassen der Studienteilnehmer	105
6.6	Schwierigkeit des Produktvergleichs für beide Gruppen basierend auf einer 7-Punkte Likert Skala	106
6.7	Boxplot für die Entscheidungsdauer der beiden Gruppen	107
6.8	Entscheidungssicherheit der beiden Gruppen basierend auf einer 7-Punkte Likert Skala	110
6.9	Antworten der Teilnehmer zur Bedeutung von Kundenrezensionen basierend auf einer 7-Punkte Likert Skala	113
6.10	Produktgruppenwissen der Teilnehmer hinsichtlich der Produktgruppe „Full HD Beamer“ basierend auf einer 7-Punkte Likert Skala	114
6.11	Antworten der Teilnehmer bezüglich Bedienbarkeit und Nützlichkeit von Rated Tags	114
7.1	Benutzer-generierte Erzeugung von ähnlichen Aspekten durch verschiedenartige Schlagwörter	122
7.2	Überführung Benutzer-generierter Tags in einheitliche Produktaspekte	123
7.3	Einordnung des Klassifikationsproblems in verwandte Themengebiete	126
7.4	Schritte des KDD-Prozesses	130
7.5	Prozessübersicht zur Erstellung eines Klassifikators	132
7.6	Bestandteile und Prozessschritte innerhalb des Klassifikationsmodells	134
7.7	Vereinfachtes Beispiel zur Verdeutlichung des SVM-Klassifikators	141

7.8	Beispielhafte Darstellung der vergebenen Pro- und Kontra-Tags eines Benutzers	155
7.9	Anzahl der extrahierten Rezensionen sowie der Anteil der darin enthaltenen verschiedenartigen Tags	161
7.10	Erkennung eines Long-Tails bei der Häufigkeitsverteilung der Tags für alle Produktkategorien	162
7.11	Durchschnittliche F_1 -Werte der Hauptklassifikatoren mit und ohne Verwendung des Ensemble-Klassifikators	164
8.1	Konzeptuelles Modell der Dienst-orientierten Bereitstellung des Rated Tags-Ansatzes	172
8.2	Vier Ebenen des SaaS-Reifegradmodells	174
8.3	Angepasstes und erweitertes ER-Diagramm aus Abbildung 4.7 zur Abgrenzung der Entitäten im Rahmen des RTaaS-Dienstes	177
8.4	Beispielhaftes Sequenzdiagramm einer letztendlichen Konsistenz innerhalb von RTaaS	179
8.5	Beispielhaftes UML-Sequenzdiagramm zur Verdeutlichung einiger Dienst-schnittstellen	182
8.6	Abstrakte Darstellung der Schnittstellenkommunikation in Form eines UML-Paketdiagramms	183
8.7	Synchrone Integration des RTaaS-Dienstes	184
8.8	Asynchrone Integration des RTaaS-Dienstes	185
8.9	Beispielhafte Zuweisung eines Elements der ProductOntology an eine Produktklasse des Dienstnehmers	187
8.10	Erweitertes ER-Diagramm aus Abbildung 8.3 zur Unterstützung einer Dienstnehmer-übergreifenden Nutzung aller Tags	188
8.11	Teilen von Tag-Informationen einzelner Produkte anhand von GTIN und der Produktklassenontologie ProductOntology.org	188
9.1	Beispielhafte Darstellung einer Vergleichsmatrix unter Verwendung von Rated Tags	191
9.2	Beispielhafter Anwendungsfall für den Einsatz von bewertbaren Hashtags in Rezensionen	192

Tabellenverzeichnis

3.1	Lageparameter für die Verteilung der Social Media Features	37
3.2	Social Media Features aus dem allgemeinen Bereich (F_A) und deren Verwendungshäufigkeit	38
3.3	Social Media Features aus dem Produkt-spezifischen Bereich (F_P) und deren Verwendungshäufigkeit	39
3.4	Durchschnittliche Anzahl an Social Media Features unterteilt nach Sortimenten	40
3.5	Zuteilung der gefundenen Social Media Features zu den Bausteinen des Social Media	41
4.1	Abgrenzung des Rated Tags-Ansatzes zu verwandten Forschungsarbeiten .	74
6.1	Identifikationsnummern und URLs zu den in der Studie verwendeten Produkte	97
6.2	Durchschnittliche Bewertungen aller Rezensionen auf einer 5-Punkte Skala hinsichtlich des Merkmals „Betriebsgeräusch“	100
6.3	Abhängige Variablen, Datenniveaus und potenzielle statistische Tests für die aufgestellten Hypothesen	104
6.4	Lageparameter für die Entscheidungsdauer der Teilnehmergruppen	107
6.5	Kontingenztafel mit den Dimensionen „Teilnehmergruppe“ und „Entscheidungsqualität“	108
6.6	Zusammenfassende Darstellung der gewählten Tests und Ergebnisse für die überprüften Hypothesen	110
6.7	Die zehn am häufigsten verwendeten Tags in der Studie (absteigend sortiert nach Häufigkeit)	112
7.1	Abgrenzung der verwandten Arbeiten aus den Bereichen Identifikation und Gruppierung von Aspekten	129
7.2	Beispielhafte Feature-Vektoren für Rated Tags-Dokumente	138
7.3	Unterschiede zwischen einem Bernoulli und Multinomial Naïve Bayes-Dokumentenmodell	139
7.4	Wahrscheinlichkeitsvektoren der Basisklassifikatoren und Aggregationsvektor $\Omega(u)$ des Ensemble-Klassifikators	153
7.5	Übersicht der extrahierten Daten	155
7.6	Anzahl der unterschiedlichen Klassen je Produktkategorie	156
7.7	Wahrheitsmatrix zur Verdeutlichung der Evaluierungsmetriken	157
7.8	Übersicht der zehn häufigsten Tags pro Kategorie	160

Tabellenverzeichnis

7.9	Übersicht der Mediane für die Anzahl der Wörter und Zeichen der verwendeten Tags	163
7.10	Evaluierungsergebnisse der Hauptklassifikatoren	165
7.11	Evaluierungsergebnisse des entworfenen Ensemble-Klassifikators	166
7.12	Vergleich der String-basierten Klassifikatoren dieser Arbeit mit der Arbeit von Islam und Inkpen	169
A.1	Liste der 115 untersuchten B2C-Onlineshops	221
B.1	Fragen und Antwortmöglichkeiten des Studienfragebogens	223
B.2	Feedback einiger Rated Tags-Teilnehmer zu Frage 11 des Fragebogens . .	224

Quelltextverzeichnis

5.1	Darstellung der Basisentitäten	81
5.2	Darstellung der DTO-Entitäten für die Schnittstellenkommunikation . . .	82
5.3	Darstellung der Server-seitigen Schnittstellen	83
5.4	Stark verkürzter Auszug einer Controller-Klasse aus dem Webserver . . .	91
5.5	Stark verkürzter Auszug einer Service-Klasse	92
8.1	Entitäten für die Kommunikation mit dem RTaaS-Dienstes	179
8.2	Schnittstellen des RTaaS-Dienstes	181
8.3	Schnittstelle des Dienstnehmers zur Bereitstellung von Kategorien	182

Abkürzungsverzeichnis

AJAX	Asynchronous JavaScript and XML	18
B2B	Business-to-Business	16
B2C	Business-to-Consumer	16
BOW	Bag of words	140
C2C	Consumer-to-Consumer	16
CDSS	Customer Decision Support System	22
CORBA	Common Object Request Broker Architecture	80
CRM	Customer-Relationship-Management	173
CSS	Cascading Style Sheets	90
DOM	Document Object Model	91
DSS	Decision Support System	22
DTO	Data Transfer Object	81
EM	Expectation Maximization	128
ER	Entity-Relationship	64
ERP	Enterprise-Resource-Planning	17
EUS	Entscheidungsunterstützungssystem	22
eWOM	electronic Word-Of-Mouth	45
FN	False negative	157
FP	False positive	157
GPS	Global Positioning System	154
GTIN	Global Trade Item Number	185
GUID	Globally Unique Identifier	179
HTML	Hypertext Markup Language	18
HTTP	Hypertext Transfer Protocol	18
ICDA	Interactive Consumer Decision Aid	25
IDA	Interactive Decision Aid	25
IDL	Interface Definition Language	80
IICT	Interactive Information Comprehension Tool	25

IIMT	Interactive Information Management Tool.....	25
IIS	Internet Information Services	90
JSON	JavaScript Object Notation	18
KDD	Knowledge discovery in databases.....	130
KMU	kleine und mittlere Unternehmen.....	11
LCS	Longest Common Subsequence.....	148
LDA	Latent Dirichlet Allocation.....	128
MVC	Model-View-Controller.....	79
NLP	Natural Language Processing	127
OASIS	Organization for the Advancement of Structured Information Standards .	77
OR	Odds Ratio.....	109
ORM	Object-Relational Mapping	90
PMI	Pointwise Mutual Information.....	127
POS	Part of speech	126
RA	Recommendation Agent.....	25
REST	Representational State Transfer	83
RIA	Rich Internet Application	17
RSS	Really Simple Syndication	30
RTaaS	Rated Tags as a Service.....	171
SaaS	Software as a Service	11
SOA	Service-Oriented Architecture	77
SOC	Service-Oriented Computing	172
SVM	Support Vector Machine	131
TAM	Technology Acceptance Model	118
TN	True negative	157
TP	True positive	157
UGC	User-Generated Content.....	9
UML	Unified Modeling Language.....	59
URI	Uniform Resource Identifier.....	186
URL	Uniform Resource Locator	39
XML	eXtensible Markup Language	18

Anhang A

Untersuchung deutscher B2C-Onlineshops

Tabelle A.1: Liste der 115 untersuchten B2C-Onlineshops (aufsteigend sortiert nach Domainname)

adobe.com/downloads/	alternate.de	amazon.de
apodiscounter.de	atu.de	baby-walz.de
bader.de	baur.de	bonprix.de
brands4friends.de	buch.de	buecher.de
buy.norton.com	c-and-a.com	channel21.de
computeruniverse.net	comtech.de	conrad.de
cyberport.de	deichmann.com	dell.de
docmorris.de	douglas.de	dress-for-less.de
drucker-guenstiger.de	druckerzubehoer.de	emp.de
esprit.de	fab.de	fernseher-guenstiger.de
fressnapf.de	frontlineshop.de	galeria-kaufhof.de
getgoods.de	globetrotter.de	goertz.de
hagebau.de	handyshop.de	hardwareversand.de
heine.de	hm.com/de	hoh.de
hornbach.de	hse24.de	ikea.com/de
imwalking.de	innova24.biz	itunes.apple.com/de
jako-o.de	kapersky.com/de	karstadt.de
kfzteile24-shop.de	kidoh.de	klingsel.de
landsend.de	libri.de	lidl.de
limango.de	logitech.com	louis.de
medion.com	medpex.de	mindfactory.de
mirapodo.de	moebel-profi.de	musicload.de
mytoys.de	nero.com/deu/store.html	notebooksbilliger.de

Tabelle A.1 (Forts.)

obi.de	office-discount.de	otto.de
pearl.de	planet-sports.de	plus.de
promarkt.de	qvc.de	real-onlineshop.de
redcoon.de	reichelt.de	reifen.com
reifendirekt.de	roller.de	rossmannversand.de
sanicare.de	schuhtempel24.de	schwab.de
sheego.de	shop.haufe.de	shop-apotheke.com
software-download.mediamarkt.de	soliver.de	sportscheck.com
staples.de	store.apple.com/de	tchibo.de
telekom.de	thalia.de	thomann.de
tomtom.com	t-online-shop.de	toysrus.de
unimall.de	vente-privee.com	viking.de
voelkner.de	walbusch.de	weltbild.de
wenz.de	westfalia.de	westwing.de
www8.hp.com/de/de	yves-rocher.de	zalando.de
zooplus.de		

Anhang B

Evaluierung des Rated Tags-Ansatzes

Tabelle B.1: Fragen und Antwortmöglichkeiten des Studienfragebogens. Die Fragen 9 bis 13 wurden nur Teilnehmern der Rated Tags-Gruppe gestellt.

#	Frage	Antwortmöglichkeiten
1	Sind Sie männlich oder weiblich?	<ul style="list-style-type: none">• Männlich• Weiblich
2	Zu welcher Altersklasse gehören Sie?	<ul style="list-style-type: none">• 18-25• 26-35• 36-45• 46-55• Älter als 55
3	Wie gut kennen Sie sich mit „Full HD Beamern“ aus?	7-Punkte Likert Skala (von „Überhaupt nicht“ bis „Sehr gut“)
4	Haben Sie schon einmal vor einem Kauf Online-Rezensionen gelesen?	<ul style="list-style-type: none">• Ja• Nein
5	Wie wichtig sind Online-Rezensionen für Ihre Entscheidungsfindung?	7-Punkte Likert Skala (von „Unwichtig“ bis „Sehr wichtig“)
6	Haben Sie bereits selbst Online-Rezensionen verfasst?	<ul style="list-style-type: none">• Ja• Nein
7	Wie schwierig war es für Sie, die Produkte hinsichtlich des gesuchten Merkmals zu vergleichen?	7-Punkte Likert Skala (von „Sehr schwer“ bis „Sehr einfach“)
8	Wie sicher sind Sie, dass Sie das richtige Produkte ausgewählt haben?	7-Punkte Likert Skala (von „Sehr unsicher“ bis „Sehr sicher“)

Tabelle B.1 (Forts.)

9	Wie einfach fanden Sie die Verwendung von Rated Tags?	7-Punkte Likert Skala (von „Sehr schwer“ bis „Sehr einfach“)
10	Wie hilfreich fanden Sie Rated Tags für Ihre Entscheidungsfindung?	7-Punkte Likert Skala (von „Nicht hilfreich“ bis „Sehr hilfreich“)
11	Hatten Sie Probleme mit Rated Tags oder haben Sie Verbesserungsvorschläge?	Freitext
12	Würden Sie selbst Rated Tags erstellen?	<ul style="list-style-type: none"> • Ja • Nein
13	Wäre für Ihre Entscheidungsfindung das Vorhandensein von Rated Tags ausreichend, d.h., könnten Sie dann auf Rezensionen verzichten?	<ul style="list-style-type: none"> • Ja • Nein

Tabelle B.2: Feedback einiger Rated Tags-Teilnehmer zu Frage 11 des Fragebogens

<p>„Durch die Sternvergabe mit einer Kommastellenangabe ist die Auswertung auch ohne lesen der Beiträge möglich um keinen großen Aufwand zu betreiben bzw. nur bei Fragen bzw. genauerer Analyse die Beiträge zu nutzen.“</p>
<p>„Die Rezensionen bei Amazon sind in entsprechende Stufen eingeteilt. D.h. man kann als Kunde die jeweiligen Bewertungen von 1 - 5 Sternen filtern und dann entsprechend die dazugehörigen Rezensionen lesen. Es wäre auch sinnvoll wenn man die Rated Tags auswählen kann und entsprechend der Rated Tags die dazugehörigen Kommentare liest.</p> <p>Beispiel:</p> <p>Man wählt als Kunde zunächst die Anzahl der Sterne aus. Für mich als Kunden sind meist die negativen Rezensionen das erste Merkmal das ich in Betracht ziehe. Anschließend kann man nochmals das jeweilige Rated Tag auswählen und dieses in 5 Stufen (Anzahl der Sterne) filtern. Dadurch sehe ich als Kunde warum ein Produkt gute bzw schlechte Bewertungen in diesem Bereich erhalten hat.“</p>
<p>„Bezug zum Text wäre hilfreich. An welcher Stelle im Text wird Bezug zum RatedTag genommen.“</p>

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Dresden, den 22. Januar 2016